

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 27, 2002, 05:48:11 ; Search time 117.02 Seconds
(without alignments)
14426.928 Million cell updates/sec

Title: US-09-840-743-5
Perfect score: 6873
Sequence: 1 gttctccgcatgactgc.....aaaaaaaaaactcgag 6873

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA: *
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq: *
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq: *
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq: *
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq: *
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq: *
6: /cgn2_6/ptodata/2/ina/backfiles1.seq: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	91.6	1.3	7218	1	US-08-232-463-14
2	73	1.1	7218	1	US-08-232-463-14
3	43.6	0.6	2207	4	US-08-956-322-3
4	43.6	0.6	2329	4	US-08-956-322-1
5	43	0.6	2140	2	US-08-314-601-1
6	43	0.6	2140	5	PCT-US95-13051-1
C 7	42.2	0.6	9636	1	US-08-323-1708-1
C 8	42.2	0.6	9636	4	US-08-954-441-1
C 9	41.4	0.6	19124	2	US-08-487-826B-13
10	40.6	0.6	289	4	US-09-007-005-17
11	40.6	0.6	289	4	US-09-244-796-17
12	40.6	0.6	1298	3	US-08-948-705-3
C 13	40.4	0.6	2447	2	US-09-014-969-14
C 14	40.2	0.6	1493	1	US-08-340-820-24
C 15	40.2	0.6	1493	1	US-08-593-535-24
C 16	40.2	0.6	1602	1	US-08-530-950-3
C 17	40.2	0.6	1602	3	US-08-888-429A-3
C 18	40.2	0.6	1602	4	US-09-149-879-3
C 19	40	0.6	6243	2	US-09-056-075-1
20	39.8	0.6	87350	3	US-08-781-891-79
C 21	39	0.6	277	4	US-09-007-005-3
C 22	39	0.6	277	4	US-09-244-796-3
C 23	39	0.6	289	4	US-09-007-005-17
C 24	39	0.6	289	4	US-09-244-796-17
25	39	0.6	306	4	US-09-461-697-203
26	39	0.6	696	4	US-09-461-697-193
27	39	0.6	699	4	US-09-461-697-191

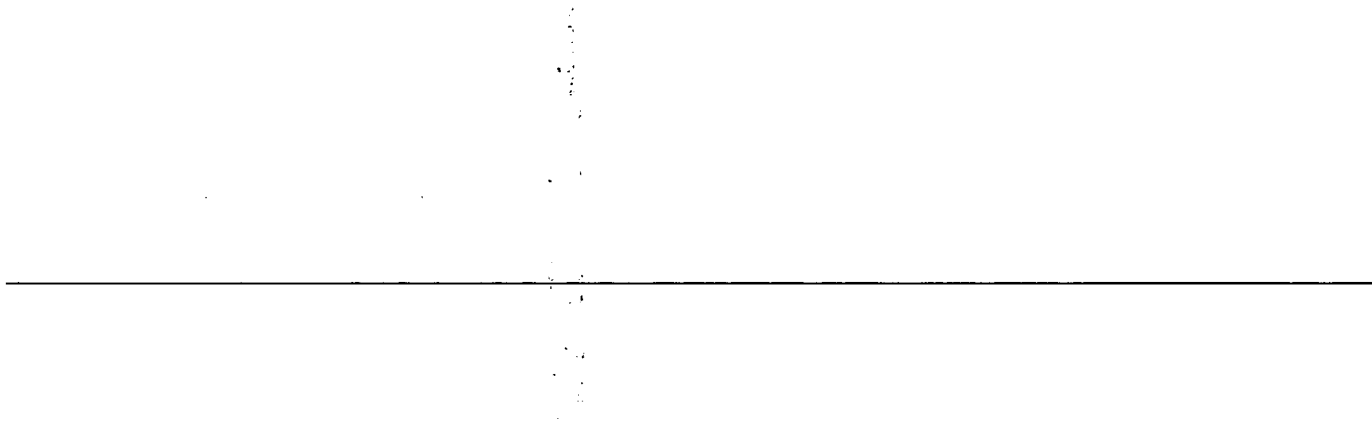
Sequence 189, App
Sequence 187, App
Sequence 185, App
Sequence 184, App
Sequence 11, Appl
Sequence 12, Appl
Sequence 6, Appl
Sequence 1, Appl
Sequence 4, Appl
Sequence 205, App
Sequence 1, Appl
Sequence 1, Appl
Sequence 33, Appl
Sequence 1, Appl
Sequence 3, Appl
Sequence 3, Appl
Sequence 1, Appl
Sequence 1, Appl

ALIGNMENTS

RESULT 1
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZ9pt-Fls
US-08-232-463-14

, MOLEC

[illegible]



THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 27, 2002, 05:47:46 ; Search time 8056.51 Seconds
(without alignments)
17852.406 Million cell updates/sec

Title: US-09-840-743-5
Perfect score: 6873
Sequence: 1 gttctccgcatgactgcgc.....aaaaaaaaaaactcgag 6873

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*

- 1: gb_ba.*
- 2: gb_htg.*
- 3: gb_in.*
- 4: gb_om.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pl.*
- 9: gb_pr.*
- 10: gb_ro.*
- 11: gb_sts.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: gb_vi.*
- 15: em_ba.*
- 16: em_fun.*
- 17: em_hum.*
- 18: em_in.*
- 19: em_mu.*
- 20: em_om.*
- 21: em_or.*
- 22: em_ov.*
- 23: em_pat.*
- 24: em_ph.*
- 25: em_pl.*
- 26: em_ro.*
- 27: em_sts.*
- 28: em_un.*
- 29: em_vi.*
- 30: em_htg_hum.*
- 31: em_htg_inv.*
- 32: em_htg_other.*
- 33: em_htgo_inv.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
------------	-------	-------------	--------	-------	-------------

1	4790.2	69.7	93045	8	ATT32M21	AL162875 Arabidops
2	3895.2	56.7	83485	8	ATTIE3	Y10157 A.thaliana
3	290	4.2	6055	8	ATSULFRED	AB051105 Euphorbia
4	182.2	2.7	361	8	AB051105	AC006919 Arabidops
5	117.8	1.7	101365	8	AC006919	AF002745 Oryza sat
6	100.4	1.5	97366	8	AP002745	I66494 Sequence 14
7	100.4	1.5	134967	8	AP002861	AF462201 Narcissus
8	91.6	1.3	7218	6	I66494	I66494 Sequence 14
9	79.4	1.2	160	6	AF462201	AL395174 T3 end of
10	73	1.1	7218	6	I66494	AL395174 T3 end of
11	66.2	1.0	863	11	CNS06EVO	AC010927 Arabidops
12	66	1.0	152988	2	AC073669	AX059537 Arabidops
13	65	0.9	5750	6	AX251465	AF128393 Arabidops
14	65	0.9	5750	6	AX346910	AL161502 Arabidops
15	63.2	0.9	685	11	CNS06EJ4	AC010927 Arabidops
16	63.2	0.9	96232	8	ATAC010927	AX059537 Arabidops
17	62	0.9	49885	6	AX059537	AF128393 Arabidops
18	62	0.9	53778	8	TLJ1	AL161502 Arabidops
19	62	0.9	198563	8	ATCHRIV14	AC010927 Arabidops
20	61.4	0.9	63365	2	AC087437	AC087437 Homo sapi
21	61.4	0.9	143585	2	AC013349	AC013349 Homo sapi
22	60.2	0.9	16033	6	AX346306	AC022851 Homo sapi
23	60.2	0.9	81120	2	AC022851	AC022851 Homo sapi
24	60.2	0.9	127709	2	CNS05TC1	AL352983 Homo sapi
25	60	0.9	222068	2	AC087188	AC087188 Homo sapi
26	59.4	0.9	146936	2	AC106157	AC106157 Rattus no
27	58.8	0.9	6145	6	AX345874	AC091325 Mus muscu
28	58.8	0.9	42612	2	AC091325	AC024285 Homo sapi
29	58.8	0.9	83440	2	AC024285	AC099851 Homo sapi
30	58.8	0.9	117193	2	AC099851	AX344566 Sequence
31	58.8	0.9	349980	6	AX344566	AX346860 Sequence
32	58.4	0.8	14006	6	AX346860	AC079333 Homo sapi
33	58.2	0.8	64612	2	AC079333	AC079417 Mus muscu
34	58.2	0.8	130540	2	AC079417	AX345169 Sequence
35	58	0.8	6419	6	AX345169	AC098448 Rattus no
36	58	0.8	169743	2	AC098448	AC084077 Homo sapi
37	57.6	0.8	56693	2	AC084077	AC015927 Homo sapi
38	57.4	0.8	131271	2	AC015927	AX346598 Sequence
39	57	0.8	6668	6	AX346598	AC023371 Homo sapi
40	57	0.8	53932	2	AC023371	AL096783 Plasmodiu
41	57	0.8	78578	2	PFMAL13P9	AC026941 Homo sapi
42	57	0.8	130808	2	AC026941	AC023185 Homo sapi
43	56.8	0.8	69284	2	AC023185	AC015830 Homo sapi
44	56.8	0.8	156550	2	AC015830	AX344562 Sequence
45	56.6	0.8	73778	6	AX344562	

ALIGNMENTS

RESULT 1
LOCUS ATT32M21 93045 bp DNA linear PLN 31-MAR-2000
DEFINITION Arabidopsis thaliana DNA chromosome 5, BAC clone T32M21 (ESSA project).
ACCESSION AL162875
VERSION AL162875.1 GI:7406444
KEYWORDS
SOURCE thale cress.
ORGANISM Arabidopsis thaliana

REFERENCE
AUTHORS Bevan, M., Terry, N., Ardiles, W., Buyssehaert, C., Dasseville, R., De Clerck, R., De Keyser, A., Neyt, P., Rouze, P., Van Den Daele, H., Villaroel, R., Gielen, J., Van Montagu, M., Bancroft, I., Mewes, H.W., Rudd, S., Lemcke, K. and Mayer, K.F.X.
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 93045)
AUTHORS EU Arabidopsis sequencing, project.
TITLE Direct Submission
JOURNAL Submitted (31-MAR-2000) MIPS, at the Max-Planck-Institut fuer

		/gene="T32M21_20"
intron	/number=9	
	complement(5430..5506)	
exon	/gene="T32M21_20"	
	/number=9	
intron	complement(5507..5573)	
	/gene="T32M21_20"	
exon	/number=10	
	complement(5574..5651)	
intron	/gene="T32M21_20"	
	/number=10	
exon	complement(5652..5783)	
	/gene="T32M21_20"	
intron	/number=11	
	complement(5784..5869)	
exon	/gene="T32M21_20"	
	/number=11	
intron	complement(5870..5947)	
	/gene="T32M21_20"	
exon	/number=12	
	complement(5948..6040)	
intron	/gene="T32M21_20"	
	/number=12	
exon	complement(6041..6151)	
	/gene="T32M21_20"	
gene	/number=13	
	complement(join(7298..7394,7484..7545,7615..7710,8136..8197,8761..8954,9380..9486,9592..9830,10135..10219))	
exon	/gene="T32M21_30"	
	complement(7298..7394)	
gene	/gene="T32M21_30"	
	/number=1	
		7298..10219
Query Match		69.7%; Score 4790.2; DB 8; Length 93045;
Best Local Similarity		92.7%; Pred. No. 0;
Matches 5229; Conservative		0; Mismatches 13; Indels 401; Gaps 3;
Qy	1	gtctccggcattgactcgctcgatgagaacacagaagcttagatcgtagcttttagctcc 60
Db	65889	GTTCPCGGCATTGACTCGCTGAGAAATCAGAAAGCTTAGATCGTGAGCTTTAGCTCC 65948
Qy	61	atttctgttattacatatttccctttttttctctctcccttttttttctctggaatt 120
Db	65949	ATTTTCTGTTTATACATATTATTCCTTTTCTCTCTCCCTTTTATCTGGAATT 66008
Qy	121	tgttctgctaaatttcagctgttacatttccgatccagagaagaatcactgggtttt 180
Db	66009	TGTTCTGCTAAATTTTCAGCTGTTTACATTTTCCGATCACCAGAGAATCACTGGGTTTT 66068
Qy	181	tatgtaatacatcatgttctctgttttctgtatcataaatctcagctattaacacctgat 240
Db	66069	TATGTTATCAATACATGTTCTGTTTCTGATCATATAATCTCAGCTATTAACACCTGAT 66128
Qy	241	tttgattctcgtaataaaaacctctgtattgctttttatcttctactttcccccataaacat 300
Db	66129	TTTGATTCTGGGTAAATAAAACCCTCTGATTGCTTTTATCTTCACTTTCCTCCCAATAACAT 66188
Qy	301	tgcttaacttattcgctcttcttttacogtttccagctaaataaattcttctacttcaat 360
Db	66189	TGCTTACTTTTATTCGCTCTCTTTTACCGTTTCCAGCTTAAATAATTTCTTCGCTATTCAAT 66248
Qy	361	gtgttctctgttttgttgatgagaaaaatactgacaaaaaatcatatttgacttttat 420
Db	66249	GTGTTTCTCGTTTGTGATGAGAAAAATATCTGNCAAAAAATCATTTATTGCAATTTTAT 66308
Qy	421	ggtgcagattcttagttaattgctgccttctcttaaccaagtcagattaaaaaggagtgctc 480
Db	66309	GGTGCAGATTCTTAGTTAATCTCGCTTCTCTTAACCAAGTCAGATTAAAAAGGAGTGTC 66368
Qy	481	gtccatgttctgtttgtttgttgagagagatttccgagagagttaggtagtcttat 540
Db	66369	GTCCATGTTGCTTTGTTTGTGTTTGGAGAGAGTTTTCGGAGAGATTAGGTGAGTGTAT 66428

Qy	541	ttggggtgagtagtgatagataaggttttgaagggggagtgattcatcaagtgtgtatgaatt 600
Db	66429	TTGGGTGAGGTAGTGATAAGGTTTGAAGGGGAGTGATTATCAAGTGTCTTATGAATT 66488
Qy	601	cgaagggtgatccgggggagatagatatattttcgagttccctttggagaatacaaaactcaacaag 660
Db	66489	CGAGGGCTGATCCGGGGGATAGATATTTTCAGTTCCTTTGGAGAAATCAAACTCAACAAG 66548
Qy	661	agttcatgggttcttgattccatttacaccacaaaaccttagatcaagtctgtatggtag 720
Db	66549	AGTTTATGGGTTCTTGGATTCCATTACACCCAAAAAACCTTAGATCAAGTCTGATGGTAG 66608
Qy	721	atgagagagtataaacaccagatctaaatgggtttccagggtgggtgaattttgtagaccagg 780
Db	66609	ATGAGAGAGTGATAAACACCAGGATCTAAATGGTTCAGGTGGTGAATTTCTAGACAGGG 66668
Qy	781	gattctgcaacactgggtgtgagatcataatgggggttttttgatcatcagggtcagggcg 840
Db	66669	GATTCTGCACACACTGGTGTGGATCATATATGGGGTTTTTGATCATGGTCTCATCAGGGCG 66728
Qy	841	ttaccaacttaagtatgatgatcaatagcttagcggatcacatgcacaaagcttggagta 900
Db	66729	TTACCAACTTAAGTATGATGATCAATAGCTTAGCGGGATCATATGCACAAGCTTGGAGTA 66788
Qy	901	atagttagagagatcttttggcaggagtgagtgacttctctctttagcaccagttatca 960
Db	66789	ATAGTGAGAGAGATCTTTTGGCAGGAGTGAGGTGACTTCTCTTTAGCACCAGTTATCA 66848
Qy	961	gaacacacacccggtaattgtagacggcgtcaatggaaatttttacttcagatcgggtatgg 1020
Db	66849	GAACACACCACCGTAATGTAGAGCCGGTCAATGGAAATTTTACTTTCAGATCGGTATGG 66908
Qy	1021	taaatgtctcttccaccagagtgccacttctcaagctcggctataatgagttgaattgg 1080
Db	66909	TAAATGGTCTTTTACCACAGAGTGGACTTCTCAAGCTGGCTATAATGAGTTGAATGG 66968
Qy	1081	atgacttgtgaatcctctgatcagatgcccttctctcctcacaaagcttgcgtgagtggtggg 1140
Db	66969	ATGACTTGTGAATCCCTGATCAGATGCCCTTCTCTTCCACAAGCTTGTGAGTGGTGGG 67028
Qy	1141	atagcttattcaagggttcgaatgtgagtgatcaaatctatttctcagtttttttttttc 1200
Db	67029	ATAGCTTATTCAAGGTTGCTCAATGTGAGTGATCAAAATCTATTTTTCAGTTTTTTTTTTC 67088
Qy	1201	ccttctccgttcttcagctacttagtagtagaagacatgaatagaatatcttaagaagt 1260
Db	67089	CCTTCTTCCGTTCTTGAGTACTTAGAGTAGAACATGAATAGAAATATCTTAGAAAAGT 67148
Qy	1261	catggttttgacagatggacctccagcgtgtaacaagcctctttacaaatttgaaattcac 1320
Db	67149	CATGGTTTGAACAGATGGACCTCCAGCGTCTAACAAAGCTCTTTACAAATTTGAATTCAC 67208
Qy	1321	caattagaagaagacagttgggttcagtcgtgaaagttcgtttcaatatatgtacogtcaa 1380
Db	67209	CAATTAGAAGAAGACAGCAGTTGGGTCAAGTCTGTAAGTTCGTTTCAATATATGACCGTCAA 67268
Qy	1381	cgccagctctttcagaacaggtgaaaagactggattccttggacagataagttacaacta 1440
Db	67269	CGCCAGCTCTTTCAGAACAGTGAAAGACTGGATTCTTGAACAGATAGTTTCAACTA 67328
Qy	1441	ctggacatgaaatcccgagacggaaatctgacaaaagtatgcagacattatggactcgt 1500
Db	67329	CTGACATGAATCCAGAGCCGAAATCTGCACAAAGTATGCAGAGCATATTGGACTCGT 67388
Qy	1501	ctgctgttaatgacaggaagctactgaacaaaatgatgycagcagacaagatgtctcgg 1560
Db	67389	CTGCTGTTAATCGCAGCGAAGCTACTGAACAAATATGATGCGCAGCAGACAAGATGTTCTGG 67448
Qy	1561	agttcgcacttaacaaaactcctcagcagaaaacccctcaaaaggaaaaggaggttcacgc 1620
Db	67449	AGTTTCGACCTTAACAAAATCTCTCAGCAAGAACCTCTCCAAAGGAAAAGGAAGTTCATGC 67508

REFERENCE	2 (bases 1 to 83485)	gene	complement(join(27911.. .28137,28504.. .28621,28713.. .30023,30153.. .30797))
AUTHORS	EU Arabidopsis sequencing, project.		/gene="TIE3_30"
TITLE	Direct Submission	CDS	complement(join(27911.. .28137,28504.. .28621,28713.. .30023,30153.. .30797))
JOURNAL	Submitted (03-APR-2000) MIPS, at the Max-Planck-Institut fuer Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail: lemcke@mips.biochem.mpg.de, mayer@mips.biochem.mpg.de Project Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK, E-mail: michael.bevan@bbsrc.ac.uk		/note="similarity to BRL protein, Homo sapiens, EMBL:AF005067"
COMMENT	Information on performance of analysis and a more detailed annotation of this entry and other sequences of chromosomes 3, 4 and 5 can be viewed at: http://www.mips.biochem.mpg.de/proj/thal/ .		/codon_start=1
FEATURES	Location/Qualifiers		/product="putative protein"
source	1. .83485		/protein_id="CAB86009.1"
	/organism="Arabidopsis thaliana"		/db_xref="GI:7413529"
	/variety="Columbia"		/translation="MPSVGMRTTRVGVVKAADGARVLRSGRIIPNVNPEKPVRAH
	/db_xref="taxon:3702"		DVDRDCSVLKNQKNGKNGKSSGKNSQPCSPKQVSEKDKVDDFPVTKRKRVRN
	/chromosome="5"		EGVGDEKTVDMKFGIVYSKRKRCLCPSSSDRSEELRSLKLYRRRKKLSORVSVLT
misc_feature	1. .25282		LVDNSCEDCWLTVEGLAMRYIRRELRLSSLASFEFLSQPINQVADHGVRLYRSP
exon	24316.. .24381		LSRGVCKFGAMSCPLFSADFVJPRMFMHMTLFLVRLPSFFFVEKSLYLNN
gene	/gene="TIE3_10"		PIEESDESALALPECTPRNGVYGLHPSVRASKTLTGNAQYRGLSGHSFQKRSS
CDS	/number=1	exon	LRRRARNLSHNAHLNNGTLPVFDISGRKNRTAAVSSKKLSRSLVLSNPSVNGIS
	/gene="TIE3_10"		IPMTKTKEELDSITCCSANLTIHSDRCTREGFSVMLEASSKSEWELVTKDCAIRYS
	/note="similarity to predicted protein, Arabidopsi thaliana"		HMAQRTMRPFSSNRITHATVMGDDNWKLEFCDQDLGFKDIYKCYERNLLEQSVK
	/codon_start=1		VPIPGVREVCGVYAEYIDNPFSPSPVSVISVNEDEVSRAMARSTALYDMSEDESM
	/product="putative protein"		LERQNKMLNEEDDQLQORAFELMIDGFEKYHFSPADOLLDKAATIGSISITLG
	/protein_id="CAB86007.1"		ROEVVEAVHDYWLKRRKRKAPLLRIFQGHQVKKTKQLLSKPVFRKRSPKROGSQHLG
	/db_xref="GI:7413527"		KAKQTSPPMVAVKAAPDEEDDILRMEEAQVLADTKTMTAIAKRRRAQILAENADLAV
	/translation="MAGKVKEKTPCWNFKCNVGGAWVMKMWLETFCANRANLIAS		YKAMRALIAEAIKEAESREVDTTV"
intron	24382.. .24629		complement(27911.. .28137)
exon	/gene="TIE3_10"		/gene="TIE3_30"
intron	/number=1		/number=1
	/gene="TIE3_10"		27911.. .30797
	/number=1		/gene="TIE3_30"
	24630.. .24752		complement(28138.. .28503)
	/gene="TIE3_10"		/gene="TIE3_30"
	/number=2		/number=1
	24753.. .24848		complement(28504.. .28621)
	/gene="TIE3_10"		/gene="TIE3_30"
	/number=2		/number=2
	24849.. .24869		complement(28622.. .28712)
	/gene="TIE3_10"		/gene="TIE3_30"
	/number=3		/number=2
	25962.. .27500		complement(28713.. .30023)
	/gene="TIE3_20"		/gene="TIE3_30"
	/number=2		/number=3
	25962.. .27500		complement(30024.. .30152)
	/gene="TIE3_20"		/gene="TIE3_30"
	/number=3		/number=3
	27501.. .27500		complement(30153.. .30797)
	/gene="TIE3_20"		/gene="TIE3_30"
	/number=4		/number=4
	27501.. .27500		complement(36227.. .36865,36944.. .37374,37440.. .37631,37768.. .38448,38589.. .38763))
	/gene="TIE3_20"		/gene="TIE3_40"
	/note="strong similarity to cytochrome P450 77A3p, Glycine max., PIR:T05948		/note="similarity to prtected prteins, Arabidopsis thaliana
	Contains Cytochrome P450 cysteine heme-iron ligand signature AA449-458"		Contains Prokaryotic membrane lipoprotein lipid attachment site AA660-670"
	/codon_start=1		/codon_start=1
	/product="cytochrom P450-like protein"		/product="putative protein"
	/protein_id="CAB86008.1"		/protein_id="CAB86010.1"
	/db_xref="GI:7413528"		/db_xref="GI:7413530"
	/translation="MEPLISFSPSLDFTFFAIIISGVFVIITRWNSNKKRLNPPG		/translation="NAISSNYLHATLMATLRSPKFAKTKVPHADLGSVSPTKKPN
	PQGVVGNLFOFARGSKPFFEYADLKKTYGFIFLRMGTRTMIILSDATLVHEALI		EEVDPDYNIOIQRKNTNRYEYIQLSOGISQGRVVEVVKDFLNDHPDAVDWENLYE
	RGALFASPAENPTTIFSCNKFTVNAAYGVPWRSRLRNVMQNMLSRLKFEKGL		TPLLKACAGKPEIKVELLRWTPGOMLPKMSQNSYHTALTVAVSGNWEIAEALVA
	ROSAMDKLERIKSEARDNDGLIWLKNARFAAFCLILEMFCGIEMDEETIEKMDIL		KNPKLEIFPGINGQIPVYVAVENTQEMARYLYTRTPVOVLAEDGYHSTLFLNAIF
	KTVLMVDRIDDLILAPFSSKERKRALEVRQGVYGVGIERRAIRAQNPGSDK		YMLGSGFLGQATHIFGGFDLYFFFIQIDIALDFLNMRRRLVTKHQIESPIIIV
	TASSFYLDTFLDKLEKRTTPSNEELVTLCSFNLQDTTGTGAIEGIAQLIAMP		LASKDPFGIQLVQLPTLPKSHANKDKSKFRTHVKYKSIYIPLKVRKSFLLFP
	EQSRYLDEIKSTQGVDRVDKQVDFLQAFKELLRKHPPTVFSITHAVMETT		DTLMRKLKGLSKGTIDEVYRLKVMHLQAKLLIGLSEETLTGLKRSERVDEALL
	LQGYDIPAGVNVVEVLPGISDPRINWNPKEFDPPDFMLGKEDADITGISGVMIPFG		FAVRYNVDFLVMIRNSELWSTRTSSSTLFLAVFROEKVFNLLYGLDDRKL
	VGRICPGLAMATIHVHLARMVQEFENCAHPPGSEIDFAGKLEFTVMVKNPLRAMV		LIADKSDGNGVLHLAGFPSPSKLASVICAPLRMORELOWKFEVERIAPEIKERVN
	KPRI"		TEQCPLEIFAKEHQGLROEAEKWKMDTAMSCSLVAALVMTFAALITYVFAAFT
	25962.. .27500		VSGGDDNSEGPNFHFDFORFIIVSDLSLISCFACACTVPIFLGLTARISDFDLVA
	/gene="TIE3_20"		LTKMITGLSILFVSIAMLIAPLSVLITMNMKNRWIVAPTILCACLPALLFVLLQYP
	/number=1		LLKEMIFSTYKGKIFDRNMTCA"

```
exon      complement(36227. .36865)
          /gene="T1E3_40"
          /number=1
gene      complement(join(36227. .36865,36944. .37374,37440. .37631,
          37768. .38448,38589. .38763))
          /gene="T1E3_40"
          36227. .38763
          /gene="T1E3_40"
          complement(36866. .36943)
          /gene="T1E3_40"
          /number=1
exon      complement(36944. .37374)
          /gene="T1E3_40"
          /number=2
intron    complement(37375. .37439)
          /gene="T1E3_40"
          /number=2
exon      complement(37440. .37631)
          /gene="T1E3_40"
          /number=3
intron    complement(37632. .37767)
          /gene="T1E3_40"
          /number=3
exon      complement(37768. .38448)
          /gene="T1E3_40"
          /number=4
intron    complement(38449. .38588)
          /gene="T1E3_40"
          /number=4
exon      complement(38589. .38763)
          /gene="T1E3_40"
          /number=5
gene      39694. .42438
          /gene="T1E3_50"
          complement(39694. .40302)
          /gene="T1E3_50"
          /number=1

Query Match      56.7%; Score 3895.2; DB 8; Length 83485;
Best Local Similarity 91.3%; Pred. No. 0;
Matches 4334; Conservative 0; Mismatches 13; Indels 401; Gaps 3;

Qy 896 gagtaaatgtagagagatcttttgggagagtgaggtgacttctctttagcaccagt 955
    |||||
Db 1 GAGTAATAGTGAGAGAGATCTTTTGGGAGGAGTGAGGTGACTTCTCCTTTAGCACCAGT 60
,
Qy 956 tatcagaacaccacgcgttaatgtagagccgtcaatggaattttacttcagatgtgg 1015
    |||||
Db 61 TATCAGAAACACCACCGTAATGTAGCGCGGTCAATGGAATTTTACTTCAGATGTGGG 120

Qy 1016 tatgtaaatggtcttccaccagagtggaacttctcaagctgagctataatgagttga 1075
    |||||
Db 121 TATGTAAATGTCCTTTTCACCACAGTGGCAGTCTCAAGCTGGCTATATGAGTTGA 180

Qy 1076 attgtagacttgtaactctgatacagatgccctctctctcacaagcttgctgagtgg 1135
    |||||
Db 181 APTGGATGACTTGTGTAATCTCTGATCAGATGCCCTTCTCCTTCAAGCTTCTGAGTGG 240

Qy 1136 tgggtagacttattcaaggttcgtcaatgtagtgatcaaatctattttcagtttttt 1195
    |||||
Db 241 TGGGATAGCTTATTCAGGTTGTCATGTAGTGATCAAAATCTATTTCAGTTTCTTTT 300

Qy 1196 ttttcccttcttcgcttcttcagtagtagagtagaacaatgaaatcttcttaag 1255
    |||||
Db 301 TTTTCCCTTCTTCGTTCTTTCAGTACTTAGTAGTAGAACATGATAGATATCTTAAG 360

Qy 1256 aaagtcaatggtttgaacagatggacctccagcgtgtaacaagcctctttacaattttaa 1315
    |||||
Db 361 AAAGTCATGGTTTGAACAGATGGACCTCCAGCGGTGAACAAGCCTCTTACAATTTGAA 420

Qy 1316 ttccacaattagaacagagagttgggtcagctcgtgaaagttcgtttcaatatgtacc 1375
    |||||
Db 421 TTCACCAATTAGAAGAGACAGTTGGGTGTCAGTCTGTGAAAGTTTCGTTCAATATGTACC 480
```

```
Qy 1376 gtcaacgcccagctcgtttcagaacaggtgaaagactggtattccttgaacagatagttac 1435
    |||||
Db 481 GTCAACGCCAGCTGCTTTCAGAACAGGTGAAAGACTGGATTCTTGAACACATAGTTAC 540

Qy 1436 aactactggaatgaaa tccagagccgaaatctctgacaaaagtatgcagagcattatgga 1495
    |||||
Db 541 AACTACTGGACATGAAATCCCGAGAGCCGAAATCTGACAAAAGTATGCAGAGCATTTATGA 600

Qy 1496 ctcgctcgtcttaatgcagcgaagactactgaacaaatgatgcagacaagaatgt 1555
    |||||
Db 601 CTCGTCTGCTGTTAATGCGAGGAAGCTACTGAAACAAATGATGCGACGACAAGATGT 660

Qy 1556 tctggagttcgaccttaacaaaactcctcagcagaaaacccctccaaaagaaaagaagt 1615
    |||||
Db 661 TCTGCAGTTCGACCTTAACAAAACCTCTCAGCAGAAAACCTTCCAAAAGAAAAGGAAGTT 720

Qy 1616 catgcccaggtggtcgctggaggcaaaccttaaaagaaagccacgcaaacctgcagaact 1675
    |||||
Db 721 CATGCCAAGGTGTCGTGGAGGCAAACTTAAAAAGAAAGCCACGCAAACTTGCAGAACT 780

Qy 1676 tcccaagtgtcgtggaagcaaaccttaaaagaaagccacgcaaacctgcagaactcagga 1735
    |||||
Db 781 TCCCAAGTGTGTCGTGGAGGCAAACTTAAAAAGAAAGCCACGCAAAAGCTGCAACTCAGGA 840

Qy 1736 aaaagtgaatctaaagaaacccggagtgccaaaagaaaaatttgaagaatcagcaac 1795
    |||||
Db 841 AAAAGTGAATCTAAAGAAACCCGGGAGTGCCAAAAGAAAATTTTGAAGAAATCAGCAAC 900

Qy 1796 taaaaagccagccaatgttgagatatgagcaacaaaagccctgaagtcaactcaaaag 1855
    |||||
Db 901 TAAAAAGCCAGCCAATGTGGAGATATGAGCAACAAAAGCCCTGAAGTCACACTCAAAAG 960

Qy 1856 ttgcagaaaactttgaaattttgacttgagaatcctgagaatcgagagcgaagtgactc 1915
    |||||
Db 961 TTGCAGAAAAGCTTTGAAATTTTGACTTGAGAATCCTGGAGATGCGAGCAAGGTGACTC 1020

Qy 1916 tgagttcgttaattccagaacagtagtgccgcaaacctggtttcttgatcagagatgc 1975
    |||||
Db 1021 TGAGTCTGAAATTTGCCAGAACAGTAGTGGCGCAAACTCGTTTTCTGAGATCAGAGATGC 1080

Qy 1976 cattggtggaactaatgttagtttctcctgattcagtgatcaaaaaatagacaagcaaatgg 2035
    |||||
Db 1081 CATTTGGTGAACATAATGTAGTTTCTCTGGATTCACTGTCACAAATAGACAAACCAATGG 1140

Qy 2036 attggggctatgaaccagccacttgaagtgtcaatggaaaccagccagatataactatc 2095
    |||||
Db 1141 ATTGGGGCTATGAACCAAGCCACTTGAAGTGTCAATGGGAAACCAAGCCAGATAAACTATC 1200

Qy 2096 tacaggagcgaactggccagagaccacaacactgatttattgactagaaaaccagcaatg 2155
    |||||
Db 1201 TACAGGAGCGAAACTGGCCAGAGACCAACACCTGATTTATTGACTAGAAAACCAAGCAATG 1260

Qy 2156 ccagttccagtggcaacccagaaacacccagttcccaatggaaaaccacaagcttgct 2215
    |||||
Db 1261 CCAGTTCCTCAGTGGCAACCCAGAACACCCAGTTCCTCAATGGAAAAACCAACAGCTTGGCT 1320

Qy 2216 tcagatgaaaaccacacttattggctttccatttgtaaccagcaacctccgatgacct 2275
    |||||
Db 1321 TCAGATGAAAAACCAACCTTATTGGCTTTTCCATTTGTGAACAGCAACCTTCGCTGACCAT 1380

Qy 2276 aagaaccagcagcctctgtggccatgggtaatacaaacactatgtatctgtagaggaaac 2335
    |||||
Db 1381 AAGAAACCAAGCAGCCTTCTGTTGGCCATGGTAAATCAACAACCTATGTATCTGATAGGAAC 1440

Qy 2336 tccacgacctgaattagtaagtggaaacagacaactagagggtcccccaaggaacaagcg 2395
    |||||
Db 1441 TCCACGGCTGCATTTAGTAAGTGGAAACAGCAACTAGGAGGTCCCAAGGAAACAAAGCG 1500

Qy 2396 gcctatttttgaataccagagactgtttaccctcgtggaaatcagctatatgtgatacc 2455
    |||||
Db 1501 GCCTATATTTTGAATACCAAGACTTGTGTTTACCTGCTGGAATATCAGCTATATGGATCACC 1560
```

Qy 2456 tacagacatgcatacaactgttattgtcaaccggaggaggaacaaacatggactactgataaa 2515
Db 1561 TACAGACATGCATCAACTGTGTTATGTCAACCGGGAGGCAACAACATGAGTACTGTATAAA 1620
Qy 2516 aaaccagcaactcggatcattataagagccagcagccttgctactcttgattgacca 2575
Db 1621 AAACCAGCAACCTGGATCATTAATAAGAGCCACAGCCTTGCCTACTTGTGATTGACCA 1680
Qy 2576 gcaacctgcaactccaaaaggttttactcaacttgaatcagatggtagctaccagcagtc 2635
Db 1681 GCAACCTGCAACTCCAAAAGGTTTACTCACTTCAATCAGATGGTAGCTACACAGCATGTC 1740
Qy 2636 atcgctgggcttcgacactcaatctcagtcacaagttccctacaacatatctacatgtgga 2695
Db 1741 ATCGCCTGGGCTTCGACCTCAATCTCAGTCACAAGTTTCTTACAACATATCTACATGTGGA 1800
Qy 2696 atcgtttccaggattttgaatggactacaggtacatgcagagagcagggcctctgc 2755
Db 1801 ATCTGTTCCAGGATTTTGAATGGGACTACAGGTACATGCCAGAGAACGAGGGCTCTCTGC 1860
Qy 2756 atcagattctttacagcaagatatccatcaaggagaaataagtacatactttctcatgagat 2815
Db 1861 ATACGATTTCTTACAGCAAGATATCCATCAAGAGAAATAAGTACATACTTTCTCATGAGAT 1920
Qy 2816 atccaatggttaattgggtgcagaagaagcgtttacctcaaaactcttctctgccaactccaat 2875
Db 1921 ATCCAATGTTAATGGTGCACAAAGCGTTTACCTCAAAACTCTTCTCTGCCAACTCCAAT 1980
Qy 2876 tatggctaaacttcagaagccagggctcgaagacagatcatcgtgcaatggaca 2935
Db 1981 TATGGCTTAAACTTGAGGAAGCCAGGGCTCGAAGAGACAGTATCATGTGCAATGGGACA 2040
Qy 2936 gacggaagaagcatgatctaaacttagctcaacagattgctcaatcaacaagatgtggagag 2995
Db 2041 GACGGAAGAAGCATGATCTAAACTTAGCTCAACAGATTGCTCAATCACAAGATGTGGAGAG 2100
Qy 2996 acataacagcagcagctgtgtggaatatattagatgctgcagaagaaaaacgaataccagaa 3055
Db 2101 ACATAAACAGCAGCAGTGTGTGGATATTTAGATGCTGCAAGAGAAAAGCAAAATCCAGAA 2160
Qy 3056 agtaqtccaagaataattgcatggcatgccacctgaggtattagaataacagagatgacc 3115
Db 2161 AGTAGTCCAAGAAATTTGCAATGGCATGCCACCTGAGGTTATAGAAATCGAGGATGATCC 2220
Qy 3116 aactgatgggcaagaaggtgtaaaaaactgcccagcatcagtaaaagtgtactctaaagg 3175
Db 2221 AACTGATGGGCAAGAAAGGTAAAAATACTGCCAGCATCAGTAAGAGTGCATCTAAGG 2280
Qy 3176 aaactcgtctccagttaaaaagcagcagaaagggagaaatgtattgtcccaaaaaagccc 3235
Db 2281 AAACCTCGTCTCCAGTTAAAAAGACAGCAGAGAAAAGGAGAAATGTATTGTCCCAAAAACGCC 2340
Qy 3236 tgcataaaaggtcgagcaggtagaaaaaatcagtaacctccgctcgtcactgctcaga 3295
Db 2341 TGCATAAAAGGGTCGAGCAGGTAGAAAAAATCAGTACTCCGCTGCTCATGCTCTCAGA 2400
Qy 3296 gatccagctttggcaacctactctccaaagcacacctttatcaagaagcaagcctaaagg 3355
Db 2401 GATCCAGCTTTGGCAACCTACTCTCTCAAGACACACCTTTTATCAAGAACGACGCTAAAGG 2460
Qy 3356 aaaagggagaagtccatcacagatccaggaaaagcaag----- 3394
Db 2461 AAAAGGGAGAAAGTCCATACAAAGATTACGAAAGCAAGAGGTAACATAATGTATTCTACA 2520
Qy 3395 ----- 3394
Db 2521 ATCTCTGTGATATAATTTTGTGATTTTGTAGTAACTGATGTGTCAAAACGAGTCTCTTATCA 2580
Qy 3395 -----aggctccatcaggagaaactctctgtgcaggaattctattgcgaa 3437
Db 2581 CTGTGTGGTCGCTGTATAGGTCCTACAGAGAACTTCTGTGTCAGGATCTTATTTCGGAA 2640
Qy 3438 ataatctcagagatgcaaaaatctgtatctagagacaaagaaagacagagacaaat 3497

Db 2641 ATAAATTACAGGATGCAAAATCTGTATCTAGGACACAAAAGAGAGCAAGACCAAAAT 2700
Qy 3498 gcaatgctctttacaaaaggagatgggtgcactgtttccctatagagcaagaagcgaaaa 3557
Db 2701 GCAATGCTTTGTACAAAGGAGATGGTGCACCTTGTCTCCCTATGAGCAAGAACGAAAA 2760
Qy 3558 ccaagaccccaaaagtgcacattgacgatgaaacaactgcataatggaacttactgatggg 3617
Db 2761 CCAAGACCCAAAAGTTGACATTTGACGATGAAACAACCTCCATATGAACTTACTGATGGG 2820
Qy 3618 aagagagatcaaaaaagaggggatgaagagaagataaaaaaagagaagtggtgggaa 3677
Db 2821 AAAGGAGATGAAAAGAAAGGGGATGAAGAGAGGATAAAAGAAAGAGAGTGTGGGAA 2880
Qy 3678 gaagaagaagagagcttctccagaggagggctgattcccttcactcgtcgcacagcactggta 3737
Db 2881 GAAGAAAGAGAGCTTCCGAGGAAGGGCTGATTCTTTCATCGCTCGCATGCACCTGGTA 2940
Qy 3738 ca----- 3739
Db 2941 CAAGGTGAAGATCCACTTCTCTCAACTCCATTTTATTCACACAAATTAGTAGAATA 3000
Qy 3740 ----- 3739
Db 3001 CTCAAAATGATGTTTGTGTGCAAAATTTTAAATTCACATAGTTAACCATGTCAAATAA 3060
Qy 3740 ----- 3739
Db 3061 TATTCATAATGCATCTTGTGAGAACAGGTGTGCATTATTATGGTGACAGCTGAATGGTTTA 3120
Qy 3740 -----aggag 3744
Db 3121 TGTGCCATTATTTCTTTTACTGCTATAGATGACCAATTTAACTTAAACGTTTACAGAG 3180
Qy 3745 atagacgtttttccatgaaggatcggtggttgattgattcggtcattcggaattttcccta 3804
Db 3181 ATAGACGTTTTTCGCCATGGAAGGGATCGGTGGTTGATTTCGGTTCATTGGAGTTTTCTCTTA 3240
Qy 3805 cacgaatgctcggatcacctttca----- 3830
Db 3241 CACAGATGTCGAGTCACTTTTCAGGATATAGTTCGCTTAAATAAATTGAGTTCCA 3300
Qy 3831 -----agctcgtcgttcactcctagct 3854
Db 3301 AAACATAGAAATTAACCCATCGTGGTTTTTACAATGCAGCTCTGCGTTTCATGCTCTAGCT 3360
Qy 3855 gctcgattccctccaaaattaagcagcagccgagaagatgaagaagatttagaagcga 3914
Db 3361 GCTCGATTCCTCCAAAATTAAGCAGCAGCCGAGAGATGAAGGAATGTTAGAACGCTTA 3420
Qy 3915 gttgtgaagatccagaaggatgattctgaacttaaatgaattccttcctcgtggcaggaa 3974
Db 3421 GTTGTGAGATCCAGAGGATGCATCTGAACCTAAATGAATTCCTTCGTGGCAGGAA 3480
Qy 3975 aaggttcaacatccatctgcacatggaagtctcgtgggttgatagtggaatcaaaagagcag 4034
Db 3481 AAGGTTCAACATCCATCTGACATGGAAGTTTCTGGGGTTGATAGTGGATCAAAAGAGCAG 3540
Qy 4035 ctaagggagactgtcaaaactcggaaattgaagatttaatttcttagaagaagattatcaa 4094
Db 3541 CTAAGGGAGCTTCAAACTCTGGAAATTAAGATTTAAATTTCTTAGAGAAGAGATATTCAA 3600
Qy 4095 aatttagaagaggaagtattatcatcacaagattctttgtatccggcgatatttcagtcg 4154
Db 3601 AATTTAGAGAGGAGATATTATCATCACAAGATCTTTTGTATCCGGCATATTTTCAGTCG 3660
Qy 4155 tgtgggagagttggtatcctctcatgttccaaatcaagcagcagaggttccctacaaccagg 4214
Db 3661 TGTGGGAGAGTTGGATCCTTCATGTTCCAAATTCAGAACTCAGACGAGAGTTTTCTTACAAACAGG 3720
Qy 4215 tgtgaacacaaaactgctcagtggaacatcacatcagtgcaaaactggagagcccaacttg 4274

Db 3721 TGTGAAACAAAACAGTGTGAGTGGACATACAAATCAGTGCACAACTGGGAGCCCAAACTTG 3780
Qy tctgatgaatttcttccaagggaatgagagacgcatactatatgaagatctggtgat 4334
Db 3781 TCTGATGAATTTGCTTCAAGGGAATGAGAGACCGCATCTATGAAGATCTGGTGAT 3840
Qy gttcagaacaaacaaactacaaaatgcgctcagagaacacctgatcttggaaaaacaaatg 4394
Db 3841 GTTCAGAAACAAGAACTACAAATGTCGCTCAGAAGAAACCTGATCTTGAAAAACAATG 3900
Qy 4395 aattggaagactctgtctgttttgggtcagccaagaaatgataactaattgccaacaact 4454
Db 3901 AATTGGAAGACTCTGTCTGTTTGGTCAGCCAAGAAATGATCTAATTTGCCAAACAACT 3960
Qy 4455 ctttcacagactatgagcagtgctgactcgacagccacatgctactagacatagagat 4514
Db 3961 CTTTCCACAGCTATGACGATGTGCGACTCGACAGCCACATGCTACTAGACATGAGGAT 4020
Qy 4515 ttggaatgcgaagtgaaggccttgggtattcttggatgtccatctcaccagagttgac 4574
Db 4021 TTTGGAATGCAAGGTGAAGGCTTGGTTATTCTTGGATGTCCATCTCACCAAGAGTTGAC 4080
Qy 4575 agagttaagaacaaatgtaccacgaggttttttcagacaagtggaagtgttccaaga 4634
Db 4081 AGAGTAAAGAACAAATGTACCACGACAGGTTTTCAGACAAAGTGAAGTGTTCCAAGA 4140
Qy 4635 gaattcacaggtcagatcataccatcaacgcctcatgaattaccaggaatggattgtcc 4694
Db 4141 GAAATCACAGTGCAGATCATACCATCAACGCTCATGANTTACCAGGNAATGGATTGTCC 4200
Qy 4695 ggttctccaagcgcgtccaagaacacacagagcagatacccaacaataatcaacaagatgag 4754
Db 4201 GGTTCCTCAAGCGCGTCCAAGAACACACCAGGACGATACCCCAACATAATCAACAAGATGAG 4260
Qy 4755 atgaataagcatcccatatttcaaaaaaacatttttggatctgctcaactcctctgaagaa 4814
Db 4261 ATGAATAAAGCATCCATTTACAAAAACATTTTGGATCTGCTCAACTCTCTCTGAAGAA 4320
Qy 4815 tgccttaacagacagtcaggtaccacaacagacaatccagcgatgctgtctaccagagagat 4874
Db 4321 TGCCTTACAGACAGTCCAGTACCAGAACAGAACATCACGCGATGCTGTCTACCGAGAGAT 4380
Qy 4875 agaactgtgaagcgtggttgatccgctcgtagtaacaattcaagcttcacagaacatatg 4934
Db 4381 AGAACTGCTGAAGACGTGGTTGATCCGCTCAGTAAACAATTCGAAGCTTACAGAACAATATG 4440
Qy 4935 gtccaatcaaatccagcaataaagagacagcgcagttgataacaaggagacaaatgcc 4994
Db 4441 GTCGAATCAAAATCCAGCAATAAAGAGCAGCGGAGTGAATCAAGGAGACAAATGCC 4500
Qy 4995 actattttacgagatgaagggaacgccttgcgtgatgggaaaaagcctacaagccagttg 5054
Db 4501 ACTATTTACGAGAGATGAAGGAGCGCTTGCCTGATGGGAAAAAGCCTACAGCCAGTGG 4560
Qy 5055 gatagctcagaagaatgctggaagggaatgaaggagacaggaacgaaacaaaaaact 5114
Db 4561 GATAGCTCTCAGAAAAAGATGTGGAGGGAATGAAGGGAGACAGGAACGAACAAAAACAAT 4620
Qy 5115 atgattccatagactatgaagcaataagacgtgtagtatcagcagagatttctgagcct 5174
Db 4621 ATGGATTCATAGACTATGAGAGCAATAAGAGCTGTCTAGTATCAGCGAGATTTCTGAGGCT 4680
Qy 5175 atcaaggaaaggagatgaataacatgttggcgttacgaataaaggatttctctagaacgg 5234
Db 4681 ATCAAGGAAGAGGGATGAATAACATGTTGGCCGTACGGAATTAAGGTAATAACTACTAATT 4740
Qy 5235 atagttaa 5242
Db 4741 TCAGTTGA 4748

RESULT 3
ATSULFRED

LOCUS ATSULFRED 6055 bp DNA linear PLN 17-JUL-1998
DEFINITION A.thaliana gene encoding sulfite reductase.
ACCESSION Y10157
VERSION Y10157.1 GI:2584720
KEYWORDS sulfite reductase.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsiis.
REFERENCE 1 (bases 1 to 6055)
AUTHORS Bork,C., Schwenh,J.D. and Hell,R.
TITLE Isolation and characterization of a gene for assimilatory sulfite
reductase from Arabidopsis thaliana
JOURNAL Gene 212 (1), 147-153 (1998)
MEDLINE 98326320
REFERENCE 2 (bases 1 to 6055)
AUTHORS Hell,R.
DIRECT SUBMISSION
SUBMITTED (18-DEC-1996) R. Hell, Lehrstuhl Fuer
pflanzensphysiologie, Ruhr-Universitaet Bochum, Universitaetsstr.
150, 44801 Bochum, D-44801, FRG
FEATURES
source Location/Qualifiers
1..6055
/organism="Arabidopsis thaliana"
/db_species="Columbia"
/db_xref="taxon:3702"
1..2269
2542..2733
/number=1
join(2542..2733,3092..3864,4004..4299,4396..4592,
4679..4815,4899..5015,5108..5265,5383..5435)
/codon_start=1
/product="sulfite reductase"
/protein_id="CAA71239.1"
/db_xref="GI:2584721"
/db_xref="SPTREMBL:O23650"
/translation="MSSTFRAPAGAATVFTADQKIRGLDALRSSHSVFLGRYGRGG
VPPPPASSSSSPIOAVTKPETATKRSKEIIEKSNFIRVPLNEELLTEAPNVNE
SAVOLTKFHGSYOQYNNREGRGSRYSFMLTKNPKGKVPNQLYLTMDDLADFGIGTL
RLTRTFQLHGVLYKQNLKATVMSIIKNMGSTLGCAGDLNRNLAPAPVYKDYLFPA
QETADNIAALLSPQSGFYIDMWVDGQFMAEPPEVVVKARNDSHGTFNFDSPETIG
TQFLPRKRVAVTPTDNDVLLNDIGVVVSDENGEPOGFNIYVGGGMRTHRMES
TFARLAEPYGVKEDILYAVKAIVTQREHRRDRKYSRMKYLIISSMGIERFDVV
EQYQKRTPEPSRELPEWFKSYLGHWEQDGCWFCGLHVDGSRVGGIMKKTLREVLEK
YKIDVRITPNQIVLCDDIKTEWKRPIITVLAQAGLLQPEFVDPLOTAMACPAFLPCP
LAITEARGIPSLKRVANFEKVLGDDYDESVIRVTCGPNCGCRRPYMAELGLVGDP
NSTQVWLGGPNTLTQTARSPFMDKVKVHDLEKCEPLFYHWKLERQTKESFETTRMG
FEKLKELIDITYKGVSQ"
2734..3091
/number=1
3092..3864
/number=2
3865..4003
/number=2
4004..4299
/number=3
4300..4395
/number=3
4396..4592
/number=4
4593..4678
/number=4
4679..4815
/number=5
4816..4898
/number=5
4899..5015
/number=6
5016..5107
/number=6
5108..5265
/number=7

[illegible]

(ftp://arthur.epm.ornl.gov/pub/xgrail), Genefinder (Phil Green, University of Washington), Genscan (Chris Burge, http://gnomic.stanford.edu/GENSCANW.html), and NetPlantGene (http://www.cbs.dtu.dk/services/NetGene2/), searches of the complete sequence against a peptide database and plant EST databases at TIGR, and manual curation based on those analyses. Annotated genes are named to indicate the level of evidence for their annotation. Genes with similarity to other proteins are named after the database hits. Genes without significant peptide similarity but with EST similarity are named as 'unknown' proteins. Genes without protein or EST similarity, that are predicted by two or more gene prediction programs over most of their length are annotated as 'hypothetical' proteins. Genes encoding tRNAs are predicted by tRNAscan-SE (Sean Eddy, http://genome.wustl.edu/eddy/tRNAscan-SE/). Simple repeats were identified by repeatmasker (Arian Smit, http://ftp.genome.washington.edu/RM/RepeatMasker.html). Genes are numbered from the top to bottom of the chromosome.

We thank the CSHL/WashU/ABI consortium for sequencing BAC clones F6P23, F5J6, T17A5, and T13L16, the ESSA group for sequencing clone F13D4, and Scott Jackson, Jiming Jiang, Klaus Meyer, Eric Richards and Satoshi Tabata for helpful assistance. In addition, we would like to thank the TIGR Bioinformatics Department, especially Lixin Zhou, Hanif Khatak, Michael E. Heaney, Lily Fu, Feng Liang, Jeremy Peterson, Michael Holmes, and Delwood Richardson for software and database support.

This work was supported by the National Science Foundation, Department of Energy and the US Department of Agriculture.

Address all correspondence to: at@tigr.org.

FEATURES

source

```
1. 101365
   Location/Qualifiers
     /organism="Arabidopsis thaliana"
     /cultivar="Columbia"
     /db_xref="taxon:3702"
     /chromosome="II"
     complement(<1..3488)
     /note="Sequence from clone F2H17"
   join(<304..486,563..613,1357..1445,1521..1626,1770..1806,
1886..2037,2371..2794,3184..3312,3391..3536,4096..4314,
4639..4719,4850..4894,4936..5055)
   /gene="At2g36370"
   <304..>5055
   /gene="At2g36370"
   /note="F2H17.2: predicted by genscan"
   join(304..486,563..613,1357..1445,1521..1626,1770..1806,
1886..2037,2371..2794,3184..3312,3391..3536,4096..4314,
4639..4719,4850..4894,4936..5055)
   /gene="At2g36370"
   /note="hypothetical protein"
   /codon_start=1
   /protein_id="AAFI8663.1"
   /db_xref="GI:6598633"
```

misc_feature

```

misc_feature
mRNA
gene
CDS

```

```

CDS
misc_feature
repeat_region
mRNA

```

```

gene
CDS
misc_feature
repeat_region
mRNA
gene
CDS

```

```

14244..14534,14613..14696,14771..14904,14986..15213,
15298..15469,15550..15804,15894..16203)
/gene="At2g36380"
<10119..16203
/genes="At2g36380"
/note="F10l11.1: related to multi drug resistance proteins
and P-glycoproteins"
join(10119..10507,10609..10729,10984..11157,11243..11533,
11613..11703,11793..12693,12781..12941,13030..13133,
13216..13381,13466..13543,13626..13735,13825..14157,
14244..14534,14613..14696,14771..14904,14986..15213,
15298..15469,15550..15804,15894..16163)
/gene="At2g36380"
/codon_start=1
/product="putative ABC transporter"
/protein_id="AAD24623.1"
/db_xref="GI:4581139"
/translation="MIGRDEDLVRTMSGRLSGTSHRSLAGAASKSFRDVPAPPTDD
VFGSRDEEDDELVEWAALERLPTDYDLRKGLMQPMVNGKIGLEDVDVDTNLAPKEK
KILMELIKFEEDNEKFLRGLEEDRVGIEVPKIEVRYENLSVEGDVRSAGRALPT
LFNVTNLTISILGLPHLPSKKRIEILKDISGIIKPSRMTLLLPSSGSKTLLQA
LACKLDDTLQMSGRITVCGHEFEFEFPQKTCAVISQHDHLHFGMTVRESUDFSGRCUG
VCFYOLLTELSREREAGIKPDPEIDAFMKSTAIISOETSLATDVYVLLKLGIDICAD
TLVDVMRRGISGGRKRLTGMELVCPATALFMEISTGLDSTTFQICKPMRQLVH
IADTVMISLLQPAPEFFELFDDIILLSEGIYVQGSRNVLVEFFETFMKCPERKGI
ADFLQEVTSKKDQEQYNNREHPYSYVSHDFSFGNSFHAGQQLASFEFVYDKAT
HPAALVTQKYGISNKDLFKACFDREMLMKRNSFYVFTQVITILASLIAMTVYFTE
GHVLLKPLSLIESVIALTYITIGFAPSAARFROLLEYFCYNOMALSFLRFGALP
RTEVANSGGTALLVYVFGGFIISKDDIPSLTWCIYTSMPMYGOTALVINEFLDE
RWGSPNDRINATKVEGLLKSGFETPEWFCIGALLGFTVLVFNFCYIILMVL
NPLNSKATTVEEGDKDKHSGTGVLTSTSSHGPKGMVLFPQLSLAFNNVNY
YVDMPEAKQAQGVGDRQLQRLDVGGAFRPGVLTALVGVSGAGKTLTMDVLADKRTGR
YVEGINSIGYPKNQATFARVSYEQNDIHSPIHTVYESLIYSAMRLSADIDTKR
EMFEVEVMELVELKPLRNSIVGLPGVDGLSTEQRKRLTIAVELVANSPIIFMDEPTSG
LDRAAAIVNRVNRVDTGTVVCTIIHQPSIDIFESFDELMLKKGGQVIYAGTLGH
HSQKLEVEFEIEGVPKIDGYNPATMWDVTTPSMSEQMSQSDVFAQIFVNSVNRNQ
ELIKELSTPPGNDLYFRTKYQAFSTQCFKWKMYSNWRYPOYNAFLMTVVI
GLVFLGLFWOTGTGKIEKQNNFAGYAAVLFGATNAATQPAVAIERTVYREK
AAGMSAIPYAISOVAIVEIMYNTIGTVYTLIYSGIMGYDWTVKFFWFYVYMLTCV
YFTLYGMMLVALPNOYIAGICLSFSLFNWLFSGELIPRQIPIKRWYVWASPVAM
TLGIITISQVGDSDRSIVHITGVGDMSLKTLKNGFGFDYDFLPVAVVHAWILFLF
AFAYGKFLNFRQR"
complement(16662..16756)
/rpt_family="(TAAA)n"
join(<16819..16960,17413..17564,17663..17898,18091..18189,
18346..18388,18859..18918,19105..19185,19275..19391,
19607..19690,20048..20169,20315..20434,20530..20559,
20778..20888,21006..21134,21217..21320,21415..21559,
21691..22248,22333..22476)
/gene="At2g36390"
<16819..22613
/genes="At2g36390"
/note="F10l11.2"
join(16819..16960,17413..17564,17663..17898,18091..18189,
18346..18388,18859..18918,19105..19185,19275..19391,
19607..19690,20048..20169,20315..20434,20530..20559,
20778..20888,21006..21134,21217..21320,21415..21559,
21691..22248,22333..22476)
/gene="At2g36390"
/codon_start=1
/product="starch branching enzyme II"
/protein_id="AAD24644.1"
/db_xref="GI:4581160"
/translation="MVTISGVFRPHLPISKKNSLSHSNEDLRNSNAVSFSLRSDS
RSSGKFAKPYSDSDSSSATTASEKLRHQSDSSSSASDQVQSDTVDSDTQVLGN
YDQKTEEAQETTLDTQSALTSYSGISYKEDFAKMSHSDVQVGQRKIPPGDGKRI
YDIDPMLNSHRNLTDYRGYQKRLREIDNEGGELEAFSRGYETFGTRATGTYRE
WAPGAKAASLIGDFNNWNAKSDYMARNDFGVEITFPNNADGSPAIPHGSRKRVIMDT
PSGKISIPAWIKYSYQVQPGIEIPYNDPREDKYAFKPRPKPTSLRYIESHVSG
MSSTEPSIANFRDVLPRIKLGYNAVOIMAIQEHAYIYASFGYHVTFFAFSSRV
GTDDKSLDKAHLEGLVMDIVHSHASKNTLDGLDMFDGTDGQYFHSGSGRYHNM
WDSRLFNYSWGEVLYLLSNARWVLEEYKFDGFGVTSNMVTHHGLQVEFTGYNE

```

YFGYSTDDAVVYLMVLNLIHGLYPEAIVVGVEDVSGMPAFVVPVEDGGVGYDLRLHM
AVADKWEILLKKEDWQVDITFTLTNRWGEKCVVYAEHQALVGVDKTIAFWLMD
KDMYDEMAVLDROATPRVDRGIALHKMTRLTJTMGLGEGYLNEMWGEFPEWIDFRT
DOHLPGRVLTAGNNGSYDKCRREFDLGDAEYLYHGLQEDRAMQNLLEYTGWTSEH
QYISRLDGRVIVFERGNLLVFNFWHTNSYSDYRIGCSVPKYLKVLDDSDNSLFGG
FNRLDSDAEFTTSDGRHDDRPCFMYAPCTAVVYAAVDDDDDDSSSLVPLGLLPE
DV"

repeat_region complement(19524..19568)
/rpt_family="(TA)n"
mRNA complement(join(22710..23271,24081..24435,24650..24841,
24940..25347))
gene /gene="At2g36400"
complement(22710..25347)
/gene="At2g36400"
/note="Fl011.3"
CDS complement(join(22836..23271,24081..24435,24650..24841,

Query Match 1.7%; Score 117.8; DB 8; Length 101365;
Best Local Similarity 61.4%; Pred. No. 3.6e-16;
Matches 224; Conservative 0; Mismatches 137; Indels 4; Gaps 2;

Qy 5802 ggagcaccatcgatagagaaaaactgtgaaccataattgaagagcgccctcgcccggg 5861

Db 62343 GAACCAGCAAAAAGGTGCATGTGTGAACCAATCATCGAAGAGCGCTTCACCGGAG 62284

Qy 5862 caagagtgcactgaataaacgagagtgatattgaagatgcttactacaatgagagacct 5921

Db 62283 CCAGAACCGCAGAGTATCAATAGCTGACATAGAGGAGCGCTTTT---TGAGGATCCA 62227

Qy 5922 gacgagatcccaacaataaaactcaacatgaacagtttgggaatgaactctacggaacac 5981

Db 62226 GAAGAAATTCCTACCATCAGCTAAACATGATGTCATTTACCACTAAGGAGATA 62167

Qy 5982 atgaagaagaacatggagctccagaaggtgacatgtccaggcttgggttgcattcat 6041

Db 62166 ATGGAACACAAAGGAACCTCAAGACGGAACATGTCACGCGCTTTAGTTGCCTACT 62107

Qy 6042 ccaacaactctatttccaaactcccaactaaagaacatagccgtctcaggacagag 6101

Db 62106 GCTGAAACTGCTTCTCTCCATGCCTAAGCTCAAGATATACGCCAGTTAAGGACAGAA 62047

Qy 6102 caccagtgtagcggtcccgagatccacatgctctccttgatgtagtataaaagaaa 6161

Db 62046 CACCGAGTGAAGA-ATCCATTTTCTGTTCTGTTTCTTCTGCTTCTAGGTAAATAATGAAA 61988

Qy 6162 ccaga 6166

Db 61987 TCAGA 61983

RESULT 6
AP002745 97366 bp DNA linear PLN 11-JUL-2000
LOCUS Oryza sativa genomic DNA, chromosome 1, PAC clone:P0489G09,
DEFINITION complete sequence.
ACCESSION AP002745
VERSION AP002745.1 GI:9049478
KEYWORDS HTG.
SOURCE Oryza sativa (cultivar:Nipponbare) DNA, clone:P0489G09.
ORGANISM Oryza sativa

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzaceae; Oryza.

1 (bases 1 to 97366)
Sasaki, T., Matsumoto, T. and Yamamoto, K.
Oryza sativa nipponbare (GA3) genomic DNA, chromosome 1, PAC
clone:P0489G09
Published Only in DataBase (2000) In press
2 (bases 1 to 97366)
Sasaki, T., Matsumoto, T. and Yamamoto, K.
Direct Submission
Submitted (07-JUL-2000) to the DDBJ/EMBL/GenBank databases. Takuji
Sasaki, National Institute of Agrobiological Resources, Rice Genome

Research Program; Kannonдай 2-1-2, Tsukuba, Ibaraki 305-8602, Japan
(E-mail:tsasaki@abr.affrc.go.jp, URL:http://rtp.dna.affrc.go.jp/,
Tel:81-298-38-7441, Fax:81-298-38-7468)
The orientation of the sequence is from SP6 to 7 of the PAC clone.

COMMENT
FEATURES
source

1. .97366
/organism="Oryza sativa"
/cultivar="Nipponbare"
/db_xref="taxon:4530"
/chromosome="1"
/clone="P0489G09"
BASE COUNT 28305 a 20201 c 20553 g 28307 t
ORIGIN

Query Match 1.5%; Score 100.4; DB 8; Length 97366;
Best Local Similarity 55.6%; Pred. No. 3.3e-12;
Matches 257; Conservative 0; Mismatches 196; Indels 9; Gaps 3;

Qy 5663 tagtgcagaacttgctttaccggcaccagagagagagcttaacaagtgcactattcc 5722

Db 17904 TAGTGCAGAGCTCGCTCTTCTGTGACCTGAAGAGAAGAGTTTAGTTACATCTGGAACCCC 17963

Qy 5723 ggtccctcccgagtcctttctcctctgtagccatccccgatgataagaactacctcttcggt 5782

Db 17964 AATAGCTGCAGAAACCTTCCACCAGACATATATAAGTTCTTAGGCCCTGTAGTAAGTCAGCT 18023

Qy 5783 ggagaaatccctagcaagtggagcaccatcgatagagaaactgtgaaccaataatga 5842

Db 18024 TGAAG---TGGAAATTCAAAACACCTGTCACCATGCTATGAACAATGCCAGCAATCATTTGA 18080

Qy 5843 agagccgcctcgccgggc---aagagtgcactgaaataaccgagagtgatattgaaga 5899

Db 18081 GGAGCCAGCAAGCCCAACACTGNAACATGAGACAGACAGATGAAAGAGTGTGCAATAGA 18140

Qy 5900 tgcttactcaatgaggaccctgacgagatcccaacaataaaactcaacatttgacaggtt 5959

Db 18141 GGATAGTGTTCGATCATCAGAAAGAAATCCCTACTATCAAGCTTAATTTTGAGGAGTT 18200

Qy 5960 tgggaatgactctacgggaacacatggaaaga---aacatgagctccaagaagtgcacat 6016

Db 18201 TACACAGAACTGGAAGAGTTATATGCAAGCAAAATATCAATGAGATTGAAGATGCTGATAT 18260

Qy 6017 gtccagagcttggtgtgtcttgcatcccaacaactctctattccaaactcccaactaaa 6076

Db 18261 GTCAAGGCTTGTGTCGCTATACTCTGAAGTTGCTTCTATCCCACTCTTAGCTCAA 18320

Qy 6077 gaacattagccgtctcaggacagaccacgaagtgtacagct 6118

Db 18321 GAATGTCAGTCGCCTAAGGACAGACACCAAGTGTATGATCT 18362

RESULT 7
AP002861
LOCUS Oryza sativa genomic DNA, chromosome 1, PAC clone:P0665D10,
DEFINITION complete sequence.
ACCESSION AP002861
VERSION AP002861.2 GI:10800055
KEYWORDS HTG.
SOURCE Oryza sativa (cultivar:Nipponbare) DNA, clone:P0665D10.
ORGANISM Oryza sativa

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzaceae; Oryza.

1 (bases 1 to 134967)
Sasaki, T., Matsumoto, T. and Yamamoto, K.
Oryza sativa nipponbare (GA3) genomic DNA, chromosome 1, PAC
clone:P0665D10
Published Only in DataBase (2000) In press
2 (bases 1 to 134967)
Sasaki, T., Matsumoto, T. and Yamamoto, K.
Direct Submission
Submitted (30-AUG-2000) Takuji Sasaki, National Institute of

Agrobiological Resources, Rice Genome Research Program; Kannonдай 2-1-2, Tsukuba, Ibaraki 305-8602, Japan
(E-mail:tsasaki@ab.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/,
Tel:81-298-38-7441, Fax:81-298-38-7468)
On Oct 12, 2000 this sequence version replaced gi:9967270.
Genes were predicted from the integrated results of the following:
GENSCAN1.0, BLAST2.0, BLASTX2.0 as well as SplicePredictor
(October 1998 version). The genomic sequence was searched against
NCBI NonRedundant Protein database, nr
(ftp://ncbi.nlm.nih.gov/blast/db) and the cDNA sequence database at
RGP. Protein homologies of the coding regions were searched against
NCBI NonRedundant Protein database with BLASTP2.0. ESTs represent
the identified cDNA sequences using BLASTN 2.0 with the
corresponding DBJ accession no. and RGP clone ID.
A gene with identity or significant homology to a protein is
classified based on the protein name to indicate the homology level
such as same name, 'putative-' and '-like protein'. A gene without
significant homology to any protein but with EST homology (covering
almost the entire length of partial sequence) is classified as an
'unknown' protein. A gene predicted with a gene prediction program
is classified as a 'hypothetical' protein.
The orientation of the sequence is from T7 to SP6 of the PAC clone.
This sequence of P0665D10 clone has an overlap with P0515G01 (DDBJ:
AP001633) clone at the position 1 to 26401 of 5' end. The sequence
of this clone starts at the position 148672 of P515G01. This
sequence of P0665D10 clone has an overlap with P0489G09 (DDBJ:
AP002745) clone at the position 54916 to 134967 of 3' end. The
sequence of this clone ends at the position 80052 of P0489G09.
Detailed information on overlap and assembly quality together with
annotation of this entry is available at
http://rgp.dna.affrc.go.jp/GenomeSeq.html.

FEATURES

source
1. 134967
/organism="Oryza sativa"
/cultivar="Nipponbare"
/db_xref="taxon:4530"
/chromosome="1"
/clone="P0665D10"
join(4426..4444,6273)..6352,6399..6564,7311..7426)
/gene="P0665D10.1"
join(4426..4444,6273)..6352,6399..6564,7311..7426)
/gene="P0665D10.1"
/note="hypothetical protein"
/codon_start=1
/protein_id="BABI6476.1"
/db_xref="GI:10800056"
/translation="MSIHESGRAYGPPKILRFVIVFKLICSPKVELCFAKIRSQYT
TVVLSWISLTENEIGFLLRTR"
complement(8189..8695)
/gene="P0665D10.2"
complement(8189..8695)
/gene="P0665D10.2"
/note="hypothetical protein"
/codon_start=1
/protein_id="BABI6477.1"
/db_xref="GI:10800057"
/translation="MNLPSKETAEDEPSPRNPSNLSPSPQRPMLHRSRRAR
GASMAITGTPSLTCPPEPSPSTARVSNSSHPVPAADVAADAVSLPPK
PSRSPPYGNWRDPTASNPSTPSRAPRRRRPAGGQLELSPTAATMAIPAENVL
EKPPFTSV"
complement(join(9395..9695,10080..10238,10320..10495,
10711..10814,11499..11616,11688..11871,11960..12270))
/gene="P0665D10.3"
complement(join(9395..9695,10080..10238,10320..10495,
10711..10814,11499..11616,11688..11871,11960..12270))
/gene="P0665D10.3"
/note="hypothetical protein"
/codon_start=1
/protein_id="BABI6478.1"
/db_xref="GI:10800058"
/translation="MRGRAAVFAVLVFLVACTAAATAAATISRKHOHRPASGAA
KAACDPAAGSVWVDESYPYLSATCPFIRAEFCRYGRPDKEYLKYRWQPPPCST

PRFDGVALLRMWSGKKVMFVGDSLALNQYESLLCMLHAAAPNARTTVPASGKVDPVT
TARFEPWRLTMMSSSEPTVTITTTSCILVRREPLASRHDHGFARRALGRYFAQRTYAR
RRARDACIFGCAAEVSTVLEFNVTIYVLYTHLVLDIVNEKAGRVLKLDLDAQARNW
LSADLVFDSHWHMPSGPTQPDYIOEGNTVVKDMDRTEAFSAKALNTWAWWSDANLL
QMTRVFFQGISPSHYRGWDGDTASATCGOTRPYNGTAYPGPIPOQAVFALSALAG
" join(14847..15053,15199..15396,15690..15863,15919..16010,
16157..16269,16388..16812)
/gene="P0665D10.4"
join(14847..15053,15199..15396,15690..15863,15919..16010,
16157..16269,16388..16812)
/gene="P0665D10.4"
/note="hypothetical protein"
/codon_start=1
/protein_id="BABI6479.1"
/db_xref="GI:10800059"
/translation="MAYKLYGALTCLIEHRKKVDPDYCEKTKLSMKFIRLIKQCNSSSD
KEFSPLSKRLSKSTLVDEGFAPWFIDVLKPLFFHGRLPDPIDNSNVAVIICQSVT
VTRASGVSSDDDELAPPTPPHPKAGVVRPSMARVGEQLHQQOPAGAGGGGRVFA
ARGAAREAAQAGSFGMDDTDCGGGGGRLVALDAGGEGEDGDEVLAIRPREGR
RRSARLPLLLDLTWTSTRILLFILNLTMSNAAEIOLOITGVRQEDVVAESDKG
TSRADATAAAAAPPVSSKRRLRWAVVLANIVFVLGGQSGVCFESAGYTTIRGDSLM
LATVQSCGAPLTVLILLLLLFQMGSPAPPLIEPSEKRWKQSSAGAGLDRLSPSAG
SVBKRGEEK"
complement(join(17545..18164,18260..18689))
/gene="P0665D10.5"
complement(join(17545..18164,18260..18689))
/gene="P0665D10.5"
/codon_start=1
/product="putative cysteine protease"
/protein_id="BABI6480.1"
/db_xref="GI:10800060"
/translation="MAPSRPAVAIVLLCTFLAFQMAADAYGGGDDGVTTQMFEM
MARFKYPCGKEKRYGRFDRNVRFIRSRPAGNSALRVNQFADLTDFEVSTH
TGAKPPCKDAPRGVDPIWLPCCIDWRYKGAVTVDKQDGCSCWAEFAVAALLEGTLQ
IRTKLTLPLSEQLVDCDSCGCGAGHTDRAFLVAAGKGTAEAGRYEYRGKCR
ADALFNHAARIGHRVAVPGDQRLATAVARQPVTAIDASGPAFOFYGSGVPPGPC
GSGSAAAAAPTTHAVTLVGYCQDGASGKKYVAKNSWCKTWGCKYILLEKDVASP
HGTCGVAVSPFYPTV"
join(22285..22792,22869..23461)
/gene="P0665D10.6"
join(22285..22792,22869..23461)
/gene="P0665D10.6"
/codon_start=1
/product="putative cysteine protease"
/protein_id="BABI6481.1"
/db_xref="GI:10800061"
/translation="MASSKPSYLSLFLSIACLSLPLSAANPSPPPPPCKSDKE
MRFVGRMAQYKGYKASWPIEHEKRYQIMKDNSNFIGSRSETEISGSGVAFAPQVTV
DSFVGNRFGDLTPGEAEQFTGNATGGLHAAPPCPIPPDSMLPCCVDMRSSGAV
TGKVFORSASCASFANAALIEGLNKIRKTGELVLSQVYVDCDTSGGSGGRADTA
LGLVAARGGVASSEEPYTGVRGCDVGKULSGHSASLSGFRVAVPPNDERQLALAVAR
QPVTAYIDAGAREFMFYKGGVYRGPCEAERNHAVAIVGCEGFGDKYWIANKWSG
DWGEQGYVYLAKDVWMPQGTGLATSPFYPTA"
complement(join(24388..24528,24577..24696,24937..25103,
25260..25361,25832..25972,26080..26163,26253..26559,
26861..26727,27036..27168,27254..27347,27356..27496,
28444..28512,29291..29499,29598..29792,30489..30538,
30822..31230,31307..31652,32161..32267))
/gene="P0665D10.7"
complement(join(24388..24528,24577..24696,24937..25103,
25260..25361,25832..25972,26080..26163,26253..26559,
26861..26727,27036..27168,27254..27347,27356..27496,
28444..28512,29291..29499,29598..29792,30489..30538,
30822..31230,31307..31652,32161..32267))
/gene="P0665D10.7"
/note="hypothetical protein"
/codon_start=1
/protein_id="BABI6482.1"
/db_xref="GI:10800062"
/translation="MAILETDNDLLSLPGLSANTRIGESPQHSDLLGLRQLCR
RRRTSLPRRRNRQLPRLSPHSTATPTEDRAVVAVPGRSHRRPPPHSTAAPRRSQGRP


```

TITLE      Direct Submission
JOURNAL    Submitted (26-DEC-2001) Environmental Horticulture, University of
           California, One Shields Avenue, Davis, CA 95616, USA
FEATURES   Location/Qualifiers
           source
             1..160
               /organism="Narcissus pseudonarcissus"
               /cultivar="Dutch Master"
               /db_xref="taxon:39639"
               /tissue_type="4-day old flower tepal"
               <1..>160
               /codon_start=1
               /product="unknown"
               /protein_id="AA169364.1"
               /db_xref="GI:18419584"
               /translation="PSIFKGLTREGIQHCFWGRGVCVGRFDRMMRAPKPLFARLHFFA
               SKGRNRKT"
BASE COUNT 51 a 32 c 41 g 36 t
ORIGIN
Query Match      1.2%; Score 79.4; DB 8; Length 160;
Best Local Similarity 69.9%; Pred. No. 2.9e-07;
Matches 107; Conservative 0; Mismatches 46; Indels 0; Gaps 0;
Qy 6511 catcaattacagaggttttcaacgagcagatcacagtctgttttggaagatttcg 6570
Db 2 CATCAATATTTAAAGCGCTTACAACTGAAGGATACAGCATTGCTTTTGGAGAGGTTTG 61
Qy 6571 tatgttcggtgattcgaacagaagacagagaccgcgtccattatgccaaggttgc 6630
Db 62 TGTGCGTGAAGGGATTCGATCGGATGATGAGAGCACCAAAACCTCTCTTCAGAGGCTC 121
Qy 6631 attttcctgcagcaaatgaagaacacacaaa 6663
Db 122 ACTTCCAGCAGCAAGCAAGNAGTACGAAACAAAA 154

RESULT 10
166494
LOCUS      I66494
DEFINITION Sequence 14 from patent US 5670367.
ACCESSION I66494
VERSION   I66494.1 GI:2724471
KEYWORDS
SOURCE    Unknown.
ORGANISM  Unclassified.
REFERENCE 1 (bases 1 to 7218)
AUTHORS   Dörner,F., Scheiflinger,F. and Falkner,F.Gunter.
TITLE     Recombinant fowlpox virus
JOURNAL   Patent: US 5670367-A 14 23-SEP-1997;
FEATURES   Location/Qualifiers
           source
             1..7218
               /organism="unknown"
BASE COUNT 1944 a 1491 c 1486 g 1929 t 368 others
ORIGIN
Query Match      1.1%; Score 73; DB 6; Length 7218;
Best Local Similarity 3.5%; Pred. No. 6.7e-06;
Matches 13; Conservative 227; Mismatches 127; Indels 0; Gaps 0;
Qy 50 ctttagctcattctctgttattacattatttcctttttctctcctctttt 109
Db 1091 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1150
Qy 110 tatctggaattgttctgctaaaatttcacgtgttacattttccgatacagaagaat 169
Db 1151 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1210
Qy 170 cactgggttttatgttaatacaatcacatgctcctgtttcttgatcataaatacagctat 229
Db 1211 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1270
```

```

Qy 230 taacacctgatttgattcgtcgtaataaaaaacctctgattgttgcatttctcactttc 289
Db 1271 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1330
Qy 290 ccataaacattgcttacttatttcgctctcttcttaccgtttccagctaaaaattctt 349
Db 1331 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1390
Qy 350 cgcattcaatgcttctcgtttgttgatgagaaaaatatctgacaaaaaatcattta 409
Db 1391 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1450
Qy 410 ttgcatt 416
Db 1451 TCTCTTT 1457

RESULT 11
CNS06EVQ      863 bp      DNA      linear      STS 10-JAN-2001
LOCUS        T7 end of clone AR0AA025B09 of library AROAA from strain CBS 732 of
DEFINITION   Zygosaccharomyces rouxii, sequence tagged site.
ACCESSION    AL395628
VERSION      AL395628.1 GI:12147375
KEYWORDS     STS.
SOURCE       Zygosaccharomyces rouxii.
ORGANISM     Zygosaccharomyces rouxii.
REFERENCE    1 (bases 1 to 863)
AUTHORS      de Montigny,J., Straub,M., Potier,S., Tekala,F., Dujon,B.,
              Wincker,P., Artiguenave,F. and Souciet,J.
              Genomic Exploration of the Hemiascomycetous Yeasts: 8.
              Zygosaccharomyces rouxii(1)
              FEBS Lett. 487 (1), 52-55 (2000)
JOURNAL      1152883
PUBMED
REFERENCE    2 (bases 1 to 863)
AUTHORS      Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,
              Bolotin-Fukuhara,M., Bon,E., Bröttier,P., Casaregola,S.,
              de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B.,
              Malpertuy,A., Neuvéglise,C., Ozier-Kalogeropoulos,O., Potier,S.,
              Saurin,W., Tekala,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,
              Wincker,P. and Weissenbach,J.
              Genomic Exploration of the Hemiascomycetous Yeasts: 1. A set of
              yeast species for molecular evolution studies(1)
              FEBS Lett. 487 (1), 3-12 (2000)
JOURNAL      1152876
PUBMED
REFERENCE    3 (bases 1 to 863)
AUTHORS      Genoscope.
              Direct Submission
              Submitted (06-SEP-2000) Genoscope - Centre National de Sequencage,
              sequef@genoscope.cns.fr - Web : www.genoscope.cns.fr)
              This STS is part of a random genomic sequencing program of thirteen
              yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
              exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,
              Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
              lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
              angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,
              Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
              5 kb were prepared and both extremities were sequenced. See
              keywords for description of this sequence and for the sequence of
              the other extremity of this insert.
FEATURES     Location/Qualifiers
           source
             1..863
               /organism="Zygosaccharomyces rouxii"
               /strain="CBS 732"
               /db_xref="taxon:4956"
               /clone="AR0AA025B09"
               /clone_lib="AR0AA"
               /note="end : T7"
             <3..>155
               misc_feature
```

```
/note="similar to Saccharomyces cerevisiae ORF YGR184c [
UBR1 ; ubiquitin-protein ligase ]"
evidence=not_experimental
BASE COUNT   196 a 105 c 60 g 417 t 85 others
ORIGIN

Query Match      1.08; Score 66.2; DB 11; Length 863;
Best Local Similarity 42.3%; Pred. No. 0.00027;
Matches 157; Conservative 40; Mismatches 173; Indels 1; Gaps 1;

Oy  51 ttttagctccatttctgttatttattacatatatttccctttttttctctctccctttttt 110
Db  255 TTCTATTATTTATATATATATATTTTATTKTATGTCCTTTTCTTTTCTTTTCTTTT 314

Oy  111 atctgggaattgtctgtaaaattccagctgtgtacatttccgacatccagagaagatc 170
Db  315 WTTYCYCHWTTTTTAACTTYYTTTKTTCYTYAAATTTTTTWTATWYACAAACTT 374

Oy  171 actgggttttattgtaatacatatgtctctctgtttctgtatcataaaatctcagctatt 230
Db  375 CTCTCTTTWTTTWTWTTTATATTTTYYATTTTCCTTWCCTTTATPAYWTTTTTT 434

Oy  231 aacacctgatttgcgtcgtaataaaacctgtgatttgcgttttattcttccatttcc 290
Db  435 YCCAYTCAWTTTTTTCYTCWACTCACAACTTAATTTTWTWTTWYWGATTATA 494

Oy  291 ccataaacattgctacttattcgtctctct-tttaccggtttccagctaaaaatctct 349
Db  495 TTACTATTCTCTTATTATTTTWTWTTTWTWTTTAAATTTTAAACCCACTCWWATTTWT 554

Oy  350 cgtattcaatgtgttctcgttttggtagagaaaaatattcgacaaaataatcattta 409
Db  555 AWTTTTTTATTGTTTCTTAAWATTWTTACTTWTCTTCTCTWTAATTAATTTATTA 614

Oy  410 ttgcattttat 420
Db  615 TATTATTTTTT 625

RESULT 12
AC073669/c
LOCUS      AC073669      152968 bp      DNA      linear      HTG 29-JUN-2000
DEFINITION Mus musculus clone CT7-273D11, WORKING DRAFT SEQUENCE, 39 unordered
            pieces.
ACCESSION  AC073669
VERSION    AC073669.1 GI:8810286
KEYWORDS   HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE     house mouse.
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
REFERENCE  1 (bases 1 to 152968)
            DOE Joint Genome Institute.
            Sequencing of Mouse
            Unpublished
REFERENCE  2 (bases 1 to 152968)
            DOE Joint Genome Institute.
            Direct Submission
            Submitted (29-JUN-2000) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
            -----Genome Center
            Center: Joint Genome Institute
            Center Code: JGI
            Web site: http://www.jgi.doe.gov
            -----
Project Information
Center Project Name: 1058096
Center clone name: RG-MBAC_273D11
-----
Summary Statistics
Consensus quality: 118135 bases at least Q40
Consensus quality: 134236 bases at least Q30
```

```
Consensus quality: 137982 bases at least Q20
Estimated insert size: 100000; pulse field gel estimation
Estimated insert size: 149168; sum-of-contigs estimation
Quality coverage: 7.41 in Q20 bases; pulse field gel estimation
Quality coverage: 4.96 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 39 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1093: contig of 1093 bp in length
* 1094: gap of unknown length
* 1194: contig of 1002 bp in length
* 2196: gap of unknown length
* 2296: contig of 1040 bp in length
* 3336: gap of unknown length
* 3436: contig of 1438 bp in length
* 4874: gap of unknown length
* 4974: contig of 1111 bp in length
* 6085: gap of unknown length
* 6185: contig of 1558 bp in length
* 7742: gap of unknown length
* 7842: contig of 1103 bp in length
* 8945: gap of unknown length
* 8946: contig of 1075 bp in length
* 10120: gap of unknown length
* 10121: contig of 1255 bp in length
* 10221: gap of unknown length
* 11476: gap of unknown length
* 12700: contig of 1124 bp in length
* 12799: gap of unknown length
* 13987: contig of 1188 bp in length
* 14087: gap of unknown length
* 15182: contig of 1095 bp in length
* 15282: gap of unknown length
* 15283: contig of 1185 bp in length
* 16468: gap of unknown length
* 16568: contig of 1868 bp in length
* 18436: gap of unknown length
* 18536: contig of 2237 bp in length
* 20772: gap of unknown length
* 20773: contig of 1341 bp in length
* 22213: gap of unknown length
* 22314: contig of 1182 bp in length
* 23496: gap of unknown length
* 23596: contig of 1622 bp in length
* 25218: gap of unknown length
* 25318: contig of 2092 bp in length
* 27410: gap of unknown length
* 27510: contig of 1356 bp in length
* 28865: gap of unknown length
* 28966: contig of 2614 bp in length
* 31590: gap of unknown length
* 31680: contig of 2740 bp in length
* 34420: gap of unknown length
* 34520: contig of 3483 bp in length
* 38003: gap of unknown length
* 38103: contig of 4191 bp in length
* 42294: gap of unknown length
* 42394: contig of 4963 bp in length
* 47357: gap of unknown length
* 47456: contig of 4243 bp in length
* 51700: gap of unknown length
* 51799: contig of 3245 bp in length
* 55045: gap of unknown length
* 55145: contig of 4569 bp in length
* 59714: gap of unknown length
* 59814: contig of 5084 bp in length
* 64898: gap of unknown length
* 64998: contig of 4686 bp in length
* 69684: gap of unknown length
```

```
* 69784 74689: contig of 4906 bp in length
* 74789: gap of unknown length
* 74790 80906: contig of 6117 bp in length
* 80907 81006: gap of unknown length
* 81007 90922: contig of 9916 bp in length
* 90923 91022: gap of unknown length
* 91023 100745: contig of 9723 bp in length
* 100746 100845: gap of unknown length
* 100846 107914: contig of 7069 bp in length
* 107915 108014: gap of unknown length
* 108015 117904: contig of 9890 bp in length
* 117905 118004: gap of unknown length
* 118005 131952: contig of 13948 bp in length
* 131953 132052: gap of unknown length
* 132053 140157: contig of 8105 bp in length
* 140158 140257: gap of unknown length
* 140258 152968: contig of 12711 bp in length.
FEATURES             Location/Qualifiers
  source              1..152968
                    /organism="Mus musculus"
                    /db_xref="taxon:10090"
                    /clone="C77-273D11"
                    /clone_lib="CitbC77 mouse BAC library"
BASE COUNT          39470 a 32372 c 31940 g 45341 t 3845 others
ORIGIN

Query Match          1.0%; Score 66; DB 2; Length 152968;
Best Local Similarity 50.0%; Pred. No. 0.00022;
Matches 195; Conservative 0; Mismatches 190; Indels 5; Gaps 1;

Qy 3321 ccaagacacccctttatcaagaagcaagcctaagaagaaaggagagaaagtcacatacaagat 3380
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 59307 CGAAGAAGAAATAGAGAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAATGAAGAAGAA 59248

Qy 3381 tcagaaagcaagaggccatcaagagaaactctgtctcaggattctatcggaataa 3440
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 59247 GAAGAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA 59188

Qy 3441 attacaggatgcataactctgtatcaggagacaagaaagagaaagagacaaatgca 3500
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 59187 RAGAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 59128

Qy 3501 atggtctgtacaaaggagatggtgcactgttccctatgagacgaagaagcgaaccca 3560
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 59127 RAGAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 59068

Qy 3561 agaccaaagttgacatgacgatgaacaaactcgcataatggaacttactgatgggaa-- 3618
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 59067 GNAAGAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 59008

Qy 3619 ---aaggagatgaaaaaaggaggtgaagagaaggaataaaaaaagagaagtggtggg 3675
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 59007 RAGAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 58948

Qy 3676 aagaagaagaagagctctccgaggaagg 3705
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 58947 AAAGAGAAAGGAGCGCTTGAGTTGAAGGG 58918

RESULT 13
AX251465
LOCUS              AX251465              5750 bp      DNA      linear      PAT 05-OCT-2001
DEFINITION        Sequence 433 from Patent WO0168912.
ACCESSION         AX251465
VERSION           AX251465.1 GI:15984888
KEYWORDS          .
SOURCE            synthetic construct.
ORGANISM          synthetic construct
                  artificial sequence.
REFERENCE         1 (bases 1 to 5750)
AUTHORS          Olek, A., Piepenbrock, C. and Berlin, K.
TITLE            Diagnosis of diseases associated with tumor suppressor genes and
```

```
Oncogenes
Patent: WO 0168912-A 433 20-SEP-2001;
Epigenomics AG (DE)
FEATURES             Location/Qualifiers
  source              1..5750
                    /organism="synthetic construct"
                    /db_xref="taxon:32630"
                    /note="chemically treated genomic DNA (Homo sapiens)"
BASE COUNT          1361 a 103 c 1261 g 3025 t
ORIGIN

Query Match          0.9%; Score 65; DB 6; Length 5750;
Best Local Similarity 46.5%; Pred. No. 0.00045;
Matches 247; Conservative 0; Mismatches 280; Indels 4; Gaps 1;

Qy 65 tctgtttattacataattatttccttttttttttttttttttttttttttttttttttttttttt 124
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3627 TATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 3686

Qy 125 ctgctaaaatttccagagctgttacattttccgataccagagaagaatcacctgggtttttatg 184
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3687 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 3746

Qy 185 ttaatcaaacatgtctctgtttttctctgatacaataatcagatcatttaaacactgattttg 244
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3747 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 3806

Qy 245 attctgcgtaataaaaacctctgatttcttttttttttttttttttttttttttttttttttt 304
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3807 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 3866

Qy 305 tactttatcgtctctcttttttttttttttttttttttttttttttttttttttttttttttt 364
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3867 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 3926

Qy 365 tctcgtttgttgatgagaaaaatctctgacaaaaaatcattttatgttttttttttttttttt 424
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3927 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 3986

Qy 425 cagattcttagttaatgcgcctctctctcaacaaagtcagattaaagagagtttcgctcc 484
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3987 TTTAGGTTTAAGCGATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATCG 4046

Qy 485 atgttgctgtgtttgtgttggttgagagagtttttcgagagagttaggtgagtttttttttt 544
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4047 TTTATTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTAT 4102

Qy 545 ggtgaggtagtgataaggtttgaaaggggggagtgattcatcaagtggtttat 595
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4103 GATAGTTTAGTGTAAATTTTGTAGTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 4153

RESULT 14
AX346910
LOCUS              AX346910              5750 bp      DNA      linear      PAT 01-FEB-2002
DEFINITION        Sequence 1981 from Patent WO0200928.
ACCESSION         AX346910
VERSION           AX346910.1 GI:18494796
KEYWORDS          .
SOURCE            synthetic construct.
ORGANISM          synthetic construct
                  artificial sequence.
REFERENCE         1 (sites)
AUTHORS          Olek, A., Piepenbrock, C. and Berlin, K.
TITLE            Diagnosis of diseases associated with the immune system
JOURNAL           Patent: WO 0200928-A 1981 03-JAN-2002;
Epigenomics AG (DE)
FEATURES             Location/Qualifiers
  source              1..5750
                    /organism="synthetic construct"
                    /db_xref="taxon:32630"
                    /note="chemically treated genomic DNA (Homo sapiens)"
```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 27, 2002, 05:51:56 ; Search time 650.38 Seconds
(without alignments)
18143.782 Million cell updates/sec

Title: US-09-840-743-5
Perfect score: 6873
Sequence: 1 gtctccgcatgactgcgc.....aaaaaaaaaaactcgcg 6873

Scoring table:
IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Genesl_032802.*

- 1: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1980.DAT.*
- 2: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1981.DAT.*
- 3: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1982.DAT.*
- 4: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1983.DAT.*
- 5: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1984.DAT.*
- 6: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1985.DAT.*
- 7: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1986.DAT.*
- 8: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1987.DAT.*
- 9: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1988.DAT.*
- 10: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1989.DAT.*
- 11: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1990.DAT.*
- 12: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1991.DAT.*
- 13: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1992.DAT.*
- 14: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1993.DAT.*
- 15: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1994.DAT.*
- 16: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1995.DAT.*
- 17: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1996.DAT.*
- 18: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1997.DAT.*
- 19: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1998.DAT.*
- 20: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1999.DAT.*
- 21: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2000.DAT.*
- 22: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT.*
- 23: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001B.DAT.*
- 24: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	6873	100.0	6873	24	Arabidopsis cDNA e
2	4790.2	69.7	10620	24	Arabidopsis DMT1 (
3	4788.6	69.7	12785	24	Arabidopsis DNA fo
4	1478	21.5	1478	24	Arabidopsis DMT 5'
5	411.8	6.0	758	24	Tomato DMT related
6	352.6	5.1	638	24	Soybean DMT1 DNA.
7	351.6	5.1	657	24	Alfalfa DMT relate
8	338.8	4.9	766	24	Cotton DMT related
9	329	4.8	1134	24	Soybean DMT2 DNA.

10	253	3.7	616	24	AAS96704	Soybean DMT3 DNA.
11	233	3.4	798	24	AAS96696	Corn DMT1 DNA. 2e
12	228.2	3.3	517	24	AAS96709	Corn DMT related c
13	219.6	3.2	583	24	AAS96711	Soybean DMT relate
14	205	3.0	205	24	AAS96688	Arabidopsis DMT ge
15	191.8	2.8	595	24	AAS96714	Corn DMT related c
16	179.8	2.6	640	24	AAS96715	Corn DMT related c
17	154	2.2	557	24	AAS96698	Wheat DMT1 DNA. T
18	151	2.2	439	24	AAS96699	Wheat DMT2 DNA. T
19	144.6	2.1	488	24	AAS96718	Corn DMT related c
20	142.4	2.1	422	24	AAS96717	Corn DMT related c
21	135.6	2.0	506	24	AAS96697	Corn DMT2 DNA. 2e
22	131.8	1.9	457	24	AAS96705	Soybean DMT4 DNA.
23	121.6	1.8	624	24	AAS96700	Wheat DMT3 DNA. T
24	118	1.7	570	24	AAS96719	Corn DMT related c
c	117.8	1.7	8880	24	AAS96692	Arabidopsis DMT2 (
c	116.4	1.7	613	24	AAS96720	Corn DMT related c
26	108.4	1.5	561	24	AAS96716	Corn DMT related c
27	100.4	1.5	492	24	AAS96701	Wheat DMT4 DNA. T
28	100.4	1.5	12120	24	AAS96695	Rice DMT1 DNA. Or
29	68.4	1.0	589	24	AAS96713	Tomato DMT related
30	65	0.9	5750	22	AAS46708	Tumour suppressor
31	65	0.9	5750	22	ABL34008	Human immune syste
32	63.2	0.9	8760	24	AAS96693	Arabidopsis DMT3 (
33	60.2	0.9	16033	24	ABL33404	Human immune syste
34	58.8	0.9	6145	24	ABL32972	Human immune syste
35	58.4	0.8	14006	24	ABL33958	Human immune syste
36	58	0.8	6419	24	ABL32267	Human immune syste
37	57	0.8	6668	24	ABL33696	Human immune syste
38	55.8	0.8	7442	22	AAS46686	Tumour suppressor
39	55.8	0.8	8580	24	AAS96694	Arabidopsis DMT4 (
40	55.4	0.8	12138	24	ABL33628	Human immune syste
c	53.8	0.8	6644	20	AAX33181	Base sequence of t
c	53.8	0.8	7372	20	AAX33182	Base sequence of t
c	53.8	0.8	7797	20	AAX33180	Cowpox virus bsr f
c	53.8	0.8	7996	20	AAX33184	Base sequence of t

ALIGNMENTS

RESULT 1

AAS96689

ID AAS96689 standard; cDNA: 6873 BP.

AC AAS96689;

XX

XX

DT 26-FEB-2002 (first entry)

XX

XX Arabidopsis cDNA encoding Demeter protein.

DE

DE Demeter; DMT; Atropis; ATR; 5-methylcytosine glycosylase; ss;

KW DNA demethylation; transgenic plant; transcription modulation;

KW flowering time; endosperm development; MEDEA.

XX

OS Arabidopsis Thaliana.

XX

XX WO200180626-A1.

XX

PD 01-NOV-2001.

XX

XX 23-APR-2001; 2001WO-USI3059.

XX

XX 21-APR-2000; 2000US-0553690.

XX

XX (REGC) UNIV CALIFORNIA.

PA

PI Fischer RL, Choi Y, Hannon M, Okumuro JK, Tatarinova TV;

XX

XX WPI; 2002-055307/07.

DR P-PSDB; AAU72736.

XX

PT New polynucleotide that control plant development comprising a sequence

having a specific homology to DEMETER domains A, B or C

Disclosure: Page 57-61: 109pp: English:

The invention relates to an isolated polynucleotide sequence or their complement encoding a polypeptide having a sequence at least 40% identical to DMT (DEMTET, previously known as ATROPOS (ATR)) Domain A, B or C or their combinations. Also included are an expression cassette comprising the polynucleotide or comprising a heterologous polynucleotide under the control of a promoter at least 70% identical to DMT 5' flanking sequence, DMT 3' flanking sequence or an 5' untranslated region of DMT, a host cell comprising an exogenous polynucleotide encoding a DMT-like protein and a transgenic plant comprising a polynucleotide encoding a DMT-like protein. The expression cassette is useful for modulating transcription. The method comprises introducing the cassette into a host cell preferably *Agrobacterium* by sexual cross, and selecting a host cell with modulated transcription, where the protein is capable of exhibiting at least one of the following biological activities, which include enhanced expression of the protein in a plant results in a delay in flowering time, introduction of the protein into a cell results in modulation of methylation of chromosomal DNA in the cell, reduction of expression of the protein in a plant results in enhanced endosperm development and expressing of the protein in an *Arabidopsis* leaf results in expression of the MEDEA gene. The polynucleotide is useful for detecting a nucleic acid in a sample. DEMTET is related to 5-methylcytosine glycosylases and regulates transcription of target genes by demethylation. The present sequence encodes a DMT-like protein.

Sequence 6873 BP; 2207 A; 1385 C; 1550 G; 1731 T; 0 other;

Query Match	100.0%;	Score 6873;	DB 24;	Length 6873;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 6873:	Conservative	0;	Mismatches	0;
			Indels	0;
			Gaps	0;

Qy	1	gttctcgggaattgactcgcctcgcctgagaaatcagaaaagcttagctcgttgagcttttttagctcc	60
Db	1	gttctcgggaattgactcgcctcgcctgagaaatcagaaaagcttagctcgttgagcttttttagctcc	60
Qy	61	attttctgtttattacataattattccctttttttttttctctctcccttttttattctagaatt	120
Db	61	attttctgtttattacataattattccctttttttttttctctctcccttttttattctagaatt	120
Qy	121	tgttctgctaattttccagctgttacattttccgatacagagaagaatacaactgggtttt	180
Db	121	tgttctgctaattttccagctgttacattttccgatacagagaagaatacaactgggtttt	180
Qy	181	tatgttaatacaatacactgttctcgtttttttttctgtacataaaatctcagctataaacacctgat	240
Db	181	tatgttaatacaatacactgttctcgtttttttctgtacataaaatctcagctataaacacctgat	240
Qy	241	tttgattctcgcgtaataaaaaaacctctgatttggtttttattcttcacatttcccataaaacat	300
Db	241	tttgattctcgcgtaataaaaaaacctctgatttggtttttattcttcacatttcccataaaacat	300
Qy	301	tgcctactttattcgcctctctctttaccggtttccagcttaaaaaaatctctcgtctattcaat	360
Db	301	tgcctactttattcgcctctctctttaccggtttccagcttaaaaaaatctctcgtctattcaat	360
Qy	361	gtgtttctcgtttttgttgatgagaaaaatactgcacaaaaatcattttattgcattttat	420
Db	361	gtgtttctcgtttttgttgatgagaaaaatactgcacaaaaatcattttattgcattttat	420
Qy	421	gggtcagaattcttagttaattgtcgccttctctcaaccagtcagatttaaaaaaggagtgttc	480
Db	421	gggtcagaattcttagttaattgtcgccttctctcaaccagtcagatttaaaaaaggagtgttc	480
Qy	481	gtccaatgttctttgttttttggtgttttgagagagagttttcggaagagttaggtgaagtatt	540
Db	481	gtccaatgttctttgttttttggtgttttgagagagagttttcggaagagttaggtgaagtatt	540
Qy	541	ttaggggtgaagtagtaataaagggttttaaggggagagtgatttcataccaagtggttatgaatt	600

Db 1621 ccaaggtggtcgtggaagcgaacctaaagaaagaaagccacgcaaacctgcagaaacttccca 1680
Qy 1681 aagtggtcgtggaagcgaacctaaagaaagaaagccacgcaaacctgcagaaacttccca 1740
Db 1681 aagtggtcgtggaagcgaacctaaagaaagaaagccacgcaaacctgcagaaacttccca 1740
Qy 1741 tgaatctaaagaaacccgggagtgccaaagaaagaaatcttgaagaaatcagcaactaaaa 1800
Db 1741 tgaatctaaagaaacccgggagtgccaaagaaagaaatcttgaagaaatcagcaactaaaa 1800
Qy 1801 agccagccaatggtggaagatagcaacaaagccctgaagtcacactcaaaagtgcga 1860
Db 1801 agccagccaatggtggaagatagcaacaaagccctgaagtcacactcaaaagtgcga 1860
Qy 1861 gaaaagcttgaatttgacttggaatccctggaatcgcagagcgaagtgactctgagt 1920
Db 1861 gaaaagcttgaatttgacttggaatccctggaatcgcagagcgaagtgactctgagt 1920
Qy 1921 ctgaaattgtccagaacagtagtgccgcaaaactcgtttcttgagatcagagatgccattg 1980
Db 1921 ctgaaattgtccagaacagtagtgccgcaaaactcgtttcttgagatcagagatgccattg 1980
Qy 1981 gtgaaactaatgttagtttccctgattcagtgctcacaaatagacaagaccaatgattgg 2040
Db 1981 gtgaaactaatgttagtttccctgattcagtgctcacaaatagacaagaccaatgattgg 2040
Qy 2041 gggctatgaacccagccactgaaagtgtcaatgggaacccagccagataaactatctcacg 2100
Db 2041 gggctatgaacccagccactgaaagtgtcaatgggaacccagccagataaactatctcacg 2100
Qy 2101 gagcgaactggccagagaccacaacactgatttattgactagaaaccagcaatggccagt 2160
Db 2101 gagcgaactggccagagaccacaacactgatttattgactagaaaccagcaatggccagt 2160
Qy 2161 tccagtggaacccagaaacccagttcccaatggaaacccagcaatggccagt 2220
Db 2161 tccagtggaacccagaaacccagttcccaatggaaacccagcaatggccagt 2220
Qy 2221 tgaaaacccaaacttattggtttccatttggtaaccagcaacccatgataagaa 2280
Db 2221 tgaaaacccaaacttattggtttccatttggtaaccagcaacccatgataagaa 2280
Qy 2281 accagcagccttgggtgccaatgggaatcaaacactatgtatctgataagaaactccac 2340
Db 2281 accagcagccttgggtgccaatgggaatcaaacactatgtatctgataagaaactccac 2340
Qy 2341 ggcctggaatagtagtgaagcagcaactagaggtccccaaggaacagcgccta 2400
Db 2341 ggcctggaatagtagtgaagcagcaactagaggtccccaaggaacagcgccta 2400
Qy 2401 tattttgaatcacagactgtttacactgctggaatcagctatatggatcacctcacg 2460
Db 2401 tattttgaatcacagactgtttacactgctggaatcagctatatggatcacctcacg 2460
Qy 2461 acatgcatcaactgttatgtcaaccggagggcaacaactatgataagaaactccac 2520
Db 2461 acatgcatcaactgttatgtcaaccggagggcaacaactatgataagaaactccac 2520
Qy 2521 agcaacctggatcatttaataagagccagcagccttcgttacctttgattgaccagaac 2580
Db 2521 agcaacctggatcatttaataagagccagcagccttcgttacctttgattgaccagaac 2580
Qy 2581 ctgcaactccaaaaggttttactcaactgaaatcagatggtagctaccagcatgcatcgc 2640
Db 2581 ctgcaactccaaaaggttttactcaactgaaatcagatggtagctaccagcatgcatcgc 2640
Qy 2641 ctgggcttcgacctcattctcagtcacaagtctctacaacatactcaatgtggaatctg 2700
Db 2641 ctgggcttcgacctcattctcagtcacaagtctctacaacatactcaatgtggaatctg 2700
Qy 2701 ttccagagatttgaattgggactacaggtacatgccagagaagcagggctcctcagcag 2760
Db 2701 ttccagagatttgaattgggactacaggtacatgccagagaagcagggctcctcagcag 2760

Qy 2761 attcttacagcaagatattccatcaaggaataaagtacatactttctcatgagatatcca 2820
Db 2761 attcttacagcaagatattccatcaaggaataaagtacatactttctcatgagatatcca 2820
Qy 2821 atgtaattgggtgcaagaaagcgttacctcaaaactctctctgccaactccaattatgg 2880
Db 2821 atgtaattgggtgcaagaaagcgttacctcaaaactctctctgccaactccaattatgg 2880
Qy 2881 ctaaaactggagaaagccagggctcgaagagacagtagtcatcgtgcaatggagacagcgg 2940
Db 2881 ctaaaactggagaaagccagggctcgaagagacagtagtcatcgtgcaatggagacagcgg 2940
Qy 2941 aaaagcatgattctaaacttagctcaacagattgtctcaatcaacagatgtgagagacata 3000
Db 2941 aaaagcatgattctaaacttagctcaacagattgtctcaatcaacagatgtgagagacata 3000
Qy 3001 acagcagcacgtgtgtaaatatttagatgctgcaagaaacccagaaactcagaaagttag 3060
Db 3001 acagcagcacgtgtgtaaatatttagatgctgcaagaaacccagaaactcagaaagttag 3060
Qy 3061 tccagaaaaatttgcagtgccacccctgaggttatagaatcgaggatgatccaaactg 3120
Db 3061 tccagaaaaatttgcagtgccacccctgaggttatagaatcgaggatgatccaaactg 3120
Qy 3121 atggggcaagaaaggttaaaatattgctccagcatcagtaaaaggtgcatctaaagaaact 3180
Db 3121 atggggcaagaaaggttaaaatattgctccagcatcagtaaaaggtgcatctaaagaaact 3180
Qy 3181 cgtctccagtttaaaagacagcagaaagggaaatgtattgtcccaaaaacgcctgcaa 3240
Db 3181 cgtctccagtttaaaagacagcagaaagggaaatgtattgtcccaaaaacgcctgcaa 3240
Qy 3241 aaaagggctcagcaggttagaaaaaaatcagtaacctcgcctgcctcagagatccc 3300
Db 3241 aaaagggctcagcaggttagaaaaaaatcagtaacctcgcctgcctcagagatccc 3300
Qy 3301 agctttggcaaccttactctccaaagacacccctttatcaagaaagcaagcctaaagaaag 3360
Db 3301 agctttggcaaccttactctccaaagacacccctttatcaagaaagcaagcctaaagaaag 3360
Qy 3361 ggaagaaagtcacagattcaggaagaaagcaaggttccatccaggaagaaactctgtgct 3420
Db 3361 ggaagaaagtcacagattcaggaagaaagcaaggttccatccaggaagaaactctgtgct 3420
Qy 3421 aggaattctattgcgaaataatttacaggaatgcataatctgtatctaggaagaaagaa 3480
Db 3421 aggaattctattgcgaaataatttacaggaatgcataatctgtatctaggaagaaagaa 3480
Qy 3481 gagaacagagcaaaatgcaatggtcttgcataaagagatgggtgcactgttccctatg 3540
Db 3481 gagaacagagcaaaatgcaatggtcttgcataaagagatgggtgcactgttccctatg 3540
Qy 3541 agagcaagaaacgaaacccaaagacccaaagtgtacattgacgtgaaacaaactgcata 3600
Db 3541 agagcaagaaacgaaacccaaagacccaaagtgtacattgacgtgaaacaaactgcata 3600
Qy 3601 ggaacttactgattggaaagagatgaaagaaaggggatgaaagaaagataaaaaa 3660
Db 3601 ggaacttactgattggaaagagatgaaagaaaggggatgaaagaaagataaaaaa 3660
Qy 3661 aagagaaggtgtgggaagaaagaaagagatcttccggaggaagggctgattccctcatcg 3720
Db 3661 aagagaaggtgtgggaagaaagaaagagatcttccggaggaagggctgattccctcatcg 3720
Qy 3721 ctgcagatgcactggtaacagagatagacgttttttggccatggaagggatcggtggtg 3780
Db 3721 ctgcagatgcactggtaacagagatagacgttttttggccatggaagggatcggtggtg 3780
Qy 3781 attcggatcattggagtttctctacagaaatgtctcggaatcacctttcaagctcggt 3840
Db 3781 attcggatcattggagtttctctacagaaatgtctcggaatcacctttcaagctcggt 3840

Qy	3841	tcattgtctctagctgctcgattccctccaaaaattaaagcagcagccgagaagatgaagga	3900
Dd	3841	tcattgtctctagctgctcgattccctccaaaaattaaagcagcagccgagaagatgaagga	3900
Qy	3901	atgttagaacgctagttgttgaaagatcccaagaagatgcattctgaacttaaaatgaatctc	3960
Dd	3901	atgttagaacgctagttgttgaaagatcccaagaagatgcattctgaacttaaaatgaatctc	3960
Qy	3961	cttcgtgcccagaaaaaggttcaacatccatctgacatggaagttctcggggttgatgtg	4020
Dd	3961	cttcgtgcccagaaaaaggttcaacatccatctgacatggaagttctcggggttgatgtg	4020
Qy	4021	gatacaaaagcagcgaaggttccaaactctggaattgaagattcaattctcttag	4080
Dd	4021	gatacaaaagcagcgaaggttccaaactctggaattgaagattcaattctcttag	4080
Qy	4081	agaagagattcaaaaattgaagagagaagtattattatcatcacaaagattctttgatccgg	4140
Dd	4081	agaagagattcaaaaattgaagagagaagtattattatcatcacaaagattctttgatccgg	4140
Qy	4141	cgaatttccagtcgtgaggaggttgatccctgttccattgttcaaatcacagcgcagagt	4200
Dd	4141	cgaatttccagtcgtgaggaggttgatccctgttccattgttcaaatcacagcgcagagt	4200
Qy	4201	ttcctacaaccaggtgtaaacaaaaactgtcagtggaacatcacaaatcagtcgaactg	4260
Dd	4201	ttcctacaaccaggtgtaaacaaaaactgtcagtggaacatcacaaatcagtcgaactg	4260
Qy	4261	ggagcccaactgtctgtatgaattgttcttaagggaatgagagaccgcattctatg	4320
Dd	4261	ggagcccaactgtctgtatgaattgttcttaagggaatgagagaccgcattctatg	4320
Qy	4321	aagatctgtgtatgttcagaacaagaataactacaatgtcgtccagaagaataactgac	4380
Dd	4321	aagatctgtgtatgttcagaacaagaataactacaatgtcgtccagaagaataactgac	4380
Qy	4381	ttgaaaaaaacaatgaattggaagactctgtctgttttgggtcagccagaataatgatacta	4440
Dd	4381	ttgaaaaaaacaatgaattggaagactctgtctgttttgggtcagccagaataatgatacta	4440
Qy	4441	attggcaaaacactccttcagcagctatgagcagtgctgcagctcgacgccacatgtac	4500
Dd	4441	attggcaaaacactccttcagcagctatgagcagtgctgcagctcgacgccacatgtac	4500
Qy	4501	tagacatagaggtatttggaaatcaaggtgaaagcctgtgtattcttctgtatgccatct	4560
Dd	4501	tagacatagaggtatttggaaatcaaggtgaaagcctgtgtattcttctgtatgccatct	4560
Qy	4561	caccaagagttgacagagtaaaagacaaaaatgtaccacgcaggttttccagacaaggtg	4620
Dd	4561	caccaagagttgacagagtaaaagacaaaaatgtaccacgcaggttttccagacaaggtg	4620
Qy	4621	gaagtttccaagagaattcaacaggttcagatcataccatcaacgcctcaatgaattaccag	4680
Dd	4621	gaagtttccaagagaattcaacaggttcagatcataccatcaacgcctcaatgaattaccag	4680
Qy	4681	gaatggattgtccggttctccagcgcctccagaacacccagcagataccaccaata	4740
Dd	4681	gaatggattgtccggttctccagcgcctccagaacacccagcagataccaccaata	4740
Qy	4741	atcaacaagatgagatgaataaaagcatcccatattacaataaaacatttttgatctgtca	4800
Dd	4741	atcaacaagatgagatgaataaaagcatcccatattacaataaaacatttttgatctgtca	4800
Qy	4801	actcctctgaagaatgcttcaagacaagtcagatcataccatcaacgaaacatcacggtgct	4860
Dd	4801	actcctctgaagaatgcttcaagacaagtcagatcataccatcaacgaaacatcacggtgct	4860
Qy	4861	gtctaccgagagatagaactgctgaagcgtggtgtgattcccgctcagtaacaattcaagct	4920
Dd	4861	gtctaccgagagatagaactgctgaagcgtggtgtgattcccgctcagtaacaattcaagct	4920
Qy	4921	tacagaacatattggtcgaattccaattccagcaataaaagagcagcgcaggttgaataca	4980
Dd	4921	tacagaacatattggtcgaattccaattccagcaataaaagagcagcgcaggttgaataca	4980

Dd	4921	tacagaacatattggtcgaattccaattccagcaataaaagagcagcgcaggttgaataca	4980
Qy	4981	aggagacaaatgccactattttacgagagatgaaaggagcgttgcgtgagaaaaaagc	5040
Dd	4981	aggagacaaatgccactattttacgagagatgaaaggagcgttgcgtgagaaaaaagc	5040
Qy	5041	ctacaagccagtggtgattgctcagaaaaagatgtgagggggaatgaagggagacagaac	5100
Dd	5041	ctacaagccagtggtgattgctcagaaaaagatgtgagggggaatgaagggagacagaac	5100
Qy	5101	gaacaaaaacaatatgatttccatagactatgaagcaataaagacgtgcgtagtacaag	5160
Dd	5101	gaacaaaaacaatatgatttccatagactatgaagcaataaagacgtgcgtagtacaag	5160
Qy	5161	agatttctgaggtctacaagaaaaagggatgaataacatgttggccgtacgaataaag	5220
Dd	5161	agatttctgaggtctacaagaaaaagggatgaataacatgttggccgtacgaataaag	5220
Qy	5221	atttcttagaacgtagttaaagatcatgtgtggtatcgaccttgaatggttgagagaat	5280
Dd	5221	atttcttagaacgtagttaaagatcatgtgtggtatcgaccttgaatggttgagagaat	5280
Qy	5281	ctctcctgataaagccaaggaactatctcttgagcataaaggtctctgggttttgaaaagt	5340
Dd	5281	ctctcctgataaagccaaggaactatctcttgagcataaaggtctctgggttttgaaaagt	5340
Qy	5341	ttgaatcgtgcgactcttaacactccacaatctgtttccctgttgacacgaatttg	5400
Dd	5341	ttgaatcgtgcgactcttaacactccacaatctgtttccctgttgacacgaatttg	5400
Qy	5401	gaaggtatagcagttaggtggatgggtgctctacaacccctacccgaatcactcagt	5460
Dd	5401	gaaggtatagcagttaggtggatgggtgctctacaacccctacccgaatcactcagt	5460
Qy	5461	tacacctcctggagctataccaggtcgtcagtcacatcccaaaaatttcttggccaagac	5520
Dd	5461	tacacctcctggagctataccaggtcgtcagtcacatcccaaaaatttcttggccaagac	5520
Qy	5521	tttgcaaacctcgatccaacgaacactgtatgaattacactaccaactgattacgtttgaa	5580
Dd	5521	tttgcaaacctcgatccaacgaacactgtatgaattacactaccaactgattacgtttgaa	5580
Qy	5581	aggtattttgcacaaagtagtagaccaaatgtaatgatgtccaatgagagagagtga	5640
Dd	5581	aggtattttgcacaaagtagtagaccaaatgtaatgatgtccaatgagagagagtga	5640
Qy	5641	gaactttgccagtgcttatgctagtgcacgaactgttttaccggcaccagagagagga	5700
Dd	5641	gaactttgccagtgcttatgctagtgcacgaactgttttaccggcaccagagagagga	5700
Qy	5701	gcttaacaagtgaactattccggtccctcccgagtccttccctccgttagccatcccg	5760
Dd	5701	gcttaacaagtgaactattccggtccctcccgagtccttccctccgttagccatcccg	5760
Qy	5761	tga tagaactacacctctccgttgagaaaaatccctagcaagtgggagcaccatcgaatag	5820
Dd	5761	tga tagaactacacctctccgttgagaaaaatccctagcaagtgggagcaccatcgaatag	5820
Qy	5821	aaaactgtgaaccaataattgaagagccgctcgcgcgggcaagagtgcaactgaataa	5880
Dd	5821	aaaactgtgaaccaataattgaagagccgctcgcgcgggcaagagtgcaactgaataa	5880
Qy	5881	ccgagagtgattgaagatgcttactacaatgagccctgacgagatccacacaataa	5940
Dd	5881	ccgagagtgattgaagatgcttactacaatgagccctgacgagatccacacaataa	5940
Qy	5941	aactcaacattgaaacagtttggaaatgactctacgggaacacatggaagaacactggagc	6000
Dd	5941	aactcaacattgaaacagtttggaaatgactctacgggaacacatggaagaacactggagc	6000
Qy	6001	tccaagaaggtgacatgtccaaggtttggtgtgttgcattccaacactcttcttctt	6060
Dd	6001	tccaagaaggtgacatgtccaaggtttggtgtgttgcattccaacactcttcttctt	6060

QY 301 tgcctactttattcgtctcttctttaccggttcagctgataaaaaattcttcgtctattcaat 360
Db 1749 tgcctactttattcgtctcttctttaccggttcagctgataaaaaattcttcgtctattcaat 1808
QY 361 gtgtttctcgtttgttgagaaaaaatctgcacaaaaatcattttatgattttat 420
Db 1809 gtgtttctcgtttgttgagaaaaaatctgcacaaaaatcattttatgattttat 1868
QY 421 ggtgcagattcttagttaatgtcgtctctcttaaccagtcagattaaaaaggagtgttc 480
Db 1869 ggtgcagattcttagttaatgtcgtctctcttaaccagtcagattaaaaaggagtgttc 1928
QY 481 gtccatgttctgtttgtttgtttgttgagagagttttcggagagtttagtgtgatttat 540
Db 1929 gtccatgttctgtttgtttgtttgttgagagagttttcggagagtttagtgtgatttat 1988
QY 541 ttgggtgtgaggtagtgataagggtttgagggttgagggtgattcatcaagtgtgtatgaatt 600
Db 1989 ttgggtgtgaggtagtgataagggtttgagggttgagggtgattcatcaagtgtgtatgaatt 2048
QY 601 cgagggtgatccgggggatagatatttccagttcctttggagaaatcaaaactcaacaag 660
Db 2049 cgagggtgatccgggggatagatatttccagttcctttggagaaatcaaaactcaacaag 2108
QY 661 agttcatgggttcttgattccatttacaccccaaaaacctagatcaagtctgatgtag 720
Db 2109 agttcatgggttcttgattccatttacaccccaaaaacctagatcaagtctgatgtag 2168
QY 721 atgagagagtataaacagagatctaattgggtttccaggtgtgtaatttgtagacagg 780
Db 2169 atgagagagtataaacagagatctaattgggtttccaggtgtgtaatttgtagacagg 2228
QY 781 gattctgcaacactggtgtggataataatgggttttttgatcatggtgtcctacaggcgc 840
Db 2229 gattctgcaacactggtgtggataataatgggttttttgatcatggtgtcctacaggcgc 2288
QY 841 ttaccaacttaagtatgatcatcaatagcttagcgggatacacatgcacaagcttggagta 900
Db 2289 ttaccaacttaagtatgatcatcaatagcttagcgggatacacatgcacaagcttggagta 2348
QY 901 atagttagagagatcttttggcaggagtaggtgaactctctttagcaccagttatca 960
Db 2349 atagttagagagatcttttggcaggagtaggtgaactctctttagcaccagttatca 2408
QY 961 gaaacaccacggtaattgtagagccgtcaatggaaaattttacttcagatgtgggtatgg 1020
Db 2409 gaaacaccacggtaattgtagagccgtcaatggaaaattttacttcagatgtgggtatgg 2468
QY 1021 taaatgtctttcaccagagtggcacttctcaagctggtctataatgtagtttgaattgg 1080
Db 2469 taaatgtctttcaccagagtggcacttctcaagctggtctataatgtagtttgaattgg 2528
QY 1081 atgacttgttgaatcctgatcagatgcccttctccttccacaaagcttgcgtggtggg 1140
Db 2529 atgacttgttgaatcctgatcagatgcccttctccttccacaaagcttgcgtggtggg 2588
QY 1141 atagcttattcaaggttcgtcaatgtgagtgatcaaaatctatttccagtttttttttc 1200
Db 2589 atagcttattcaaggttcgtcaatgtgagtgatcaaaatctatttccagtttttttttc 2648
QY 1201 cctttctcgtttcttcagtaacttagagtagaacatgaattagaatatcttaagaaagt 1260
Db 2649 cctttctcgtttcttcagtaacttagagtagaacatgaattagaatatcttaagaaagt 2708
QY 1261 catggttttgaacagatggacctccagcgtgtaacaagccttcttacaatttgaattcac 1320
Db 2709 catggttttgaacagatggacctccagcgtgtaacaagccttcttacaatttgaattcac 2768
QY 1321 caattagaagagaagcagttgggttcagctcgttgaaaagttcgtttccaatatgtaccgtcaa 1380
Db 2769 caattagaagagaagcagttgggttcagctcgttgaaaagttcgtttccaatatgtaccgtcaa 2828
QY 1381 cgcccagctcgttccagaaacaggttgaaaagactggattccttgaacagatagttacaacta 1440

Db 2829 cgcccagctcgttccagaaacaggtgaaaagactggattccttgaacagatagttacaacta 2888
QY 1441 ctggacatgaaatccagagccgaaatctgacaaaaatgacagacattatgagactcgt 1500
Db 2889 ctggacatgaaatccagagccgaaatctgacaaaaatgacagacattatgagactcgt 2948
QY 1501 ctgctgttaatgcgacggaagctactgaaacaaatgatggcagacacaagatgtttctgg 1560
Db 2949 ctgctgttaatgcgacggaagctactgaaacaaatgatggcagacacaagatgtttctgg 3008
QY 1561 agttcgaccttaacaaaaactcctcagcagaaacccctcaaaaggaaaaggaaagtctatgc 1620
Db 3009 agttcgaccttaacaaaaactcctcagcagaaacccctcaaaaggaaaaggaaagtctatgc 3068
QY 1621 ccaaggtgtgttggaagcgaacaccttaaaagaaagccagcaaacctcgcagaaacttccca 1680
Db 3069 ccaaggtgtgttggaagcgaacaccttaaaagaaagccagcaaacctcgcagaaacttccca 3128
QY 1681 aagtgtgtcgtgggaagcgaacaccttaaaaggaagccacgcaaaagctgcacactcaggaaaaag 1740
Db 3129 aagtgtgtcgtgggaagcgaacaccttaaaaggaagccacgcaaaagctgcacactcaggaaaaag 3188
QY 1741 tgaatctaaagaaacccgggagtgccaaaaagaaaaatttgaaagatcagcaactaaaa 1800
Db 3189 tgaatctaaagaaacccgggagtgccaaaaagaaaaatttgaaagatcagcaactaaaa 3248
QY 1801 agccagccaatgttgagatgatgacacaaagccctgaagtcaactcaaaagtgtgca 1860
Db 3249 agccagccaatgttgagatgatgacacaaagccctgaagtcaactcaaaagtgtgca 3308
QY 1861 gaaaagctttgaaatttgcacttgagaaatcctggagatcgaggaaggtgcactctgagt 1920
Db 3309 gaaaagctttgaaatttgcacttgagaaatcctggagatcgaggaaggtgcactctgagt 3368
QY 1921 ctgaaattgtccagaaacagtagtggcgcaaacctgtttcttgagatcagagatgcccattg 1980
Db 3369 ctgaaattgtccagaaacagtagtggcgcaaacctgtttcttgagatcagagatgcccattg 3428
QY 1981 gtggaaactaatgttagtttctcgtgattcagtgacaaaatagacaaagcaaatgattgg 2040
Db 3429 gtggaaactaatgttagtttctcgtgattcagtgacaaaatagacaaagcaaatgattgg 3488
QY 2041 gggctaagaaccagccacttgaagtgcattgggaaaccagccagataaactatctacag 2100
Db 3489 gggctaagaaccagccacttgaagtgcattgggaaaccagccagataaactatctacag 3548
QY 2101 gagcgaactggccagagaccaaacactgattttattgactagaaaaccagcaatggccagt 2160
Db 3549 gagcgaactggccagagaccaaacactgattttattgactagaaaaccagcaatggccagt 3608
QY 2161 tcccagtggaacccagaaacaccagttcccaatggaaaaccaacagcttggcttcaga 2220
Db 3609 tcccagtggaacccagaaacaccagttcccaatggaaaaccaacagcttggcttcaga 3668
QY 2221 tgaaaaaccaactatttggctttccatttggtaaccagcaacctcgcatgaccataagaa 2280
Db 3669 tgaaaaaccaactatttggctttccatttggtaaccagcaacctcgcatgaccataagaa 3728
QY 2281 accagcagccttcttggccatgggtaatacaacaacctatgtatctctgataggaaactccac 2340
Db 3729 accagcagccttcttggccatgggtaatacaacaacctatgtatctctgataggaaactccac 3788
QY 2341 ggctgatttagtaagtggaaacagcaactaggaggtccccaaggaaaacagcggccta 2400
Db 3789 ggctgatttagtaagtggaaacagcaactaggaggtccccaaggaaaacagcggccta 3848
QY 2401 tatttttgaatcaccaagcttgtttacactcgttgaaatcagctatatgtgatccactacag 2460
Db 3849 tatttttgaatcaccaagcttgtttacactcgttgaaatcagctatatgtgatccactacag 3908
QY 2461 acatgcatcaactgtttatgtcaaccggagggcaacaacatggactactgataaaaaacc 2520

Dd 3909 acatgcatcaacttgttktatgtcaccggaggccaacatcggaacttgactactgataaaaaacc 3968
Qy 2521 agcaacctggatcatataaagaagccagcagccttgcgtacccttggatgaccagcaac 2580
Dd 3969 agcaacctggatcatataaagaagccagcagccttgcgtacccttggatgaccagcaac 4028
Qy 2581 ctgcaactccaaaaggttttactcaacttgaatcagatggttagctaccagcatgtcatcgc 2640
Dd 4029 ctgcaactccaaaaggttttactcaacttgaatcagatggttagctaccagcatgtcatcgc 4088
Qy 2641 ctgggtctcacctcatctcagtcacagtctctacacatatctacatgtggaatctg 2700
Dd 4089 ctgggtctcacctcatctcagtcacagtctctacacatatctacatgtggaatctg 4148
Qy 2701 ttccagattttgaaaggactacaggtacatgtccagagaaagcagggctccctgcatacgc 2760
Dd 4149 ttccagattttgaaaggactacaggtacatgtccagagaaagcagggctccctgcatacgc 4208
Qy 2761 attctttacagcaagatatccatcaaggaaataagtcacatactttctcatgagatatcca 2820
Dd 4209 attctttacagcaagatatccatcaaggaaataagtcacatactttctcatgagatatcca 4268
Qy 2821 atggttaatgggtgcaagaagcgttacctcaaaactcttctgccaaactcccaattatgg 2880
Dd 4269 atggcaatgggtgcaagaagcgttacctcaaaactcttctgccaaactcccaattatgg 4328
Qy 2881 ctaaacttgaggaaagccaggggctcgaagagacagtatcatcgtgcaatgggacagacgg 2940
Dd 4329 ctaaacttgaggaaagccaggggctcgaagagacagtatcatcgtgcaatgggacagacgg 4388
Qy 2941 aaaagcatgatctaaaacttgaactgaacagattgtctcaatcacagaagtgtggagagacata 3000
Dd 4389 aaaagcatgatctaaaacttgaactgaacagattgtctcaatcacagaagtgtggagagacata 4448
Qy 3001 acagcagcagtggtggaatatttagatctgcgaagaaacgaaataaccagaagaatag 3060
Dd 4449 acagcagcagtggtggaatatttagatctgcgaagaaacgaaataaccagaagaatag 4508
Qy 3061 tccaagaaaatttgcattggcagccacctgaggttatagaaaatcgagagtgatccaactg 3120
Dd 4509 tccaagaaaatttgcattggcagccacctgaggttatagaaaatcgagagtgatccaactg 4568
Qy 3121 atggggcaagaaaaagttaaaaatactccagcatcagtaaaagtgcatctaaaggaaact 3180
Dd 4569 atggggcaagaaaaagttaaaaatactccagcatcagtaaaagtgcatctaaaggaaact 4628
Qy 3181 cgtctcagttaaaaagacagcaagaaagagaaatgtattgtcccaaaaacgcctgcaa 3240
Dd 4629 cgtctcagttaaaaagacagcaagaaagagaaatgtattgtcccaaaaacgcctgcaa 4688
Qy 3241 aaaaggctcagcaggttagaaaaaaatcagtaacctcgccctgctcatcgctcagagatcc 3300
Dd 4689 aaaaggctcagcaggttagaaaaaaatcagtaacctcgccctgctcatcgctcagagatcc 4748
Qy 3301 agctttggcaacctactctccaaagcacactttatcaagaagaacgcttaaaaggaaag 3360
Dd 4749 agctttggcaacctactctccaaagcacactttatcaagaagaacgcttaaaaggaaag 4808
Qy 3361 ggagaaagtcctacagaatttcaggaaagaaagcag- 3394
Dd 4809 ggagaaagtcctacagaatttcaggaaagaaagcag- 4868
Qy 3395 ----- 3394
Dd 4869 tlgatataattttgagatttttagtaactgatgtgtgccaaacccagctccttatacactgtt 4928
Qy 3395 ----- aggtccatcagagaaacttctgtgtcaggattcttatttcgggaaataat 3442
Dd 4929 ggtgcgttgataggtccatcaggagaactctctgtgcaggattcttatttcgggaaataat 4988
Qy 3443 ttacagagatgcaaaatctgtatcttagagagcaaaagaaaggaacaaagcaaaatgcaat 3502
Dd 4989 ttacagagatgcaaaatctgtatcttagagagcaaaagaaaggaacaaagcaaaatgcaat 5048

Qy 3503 ggtcttgcacaaagagatggtgcacttgttccctatgagcagaagcgaaaaaccaag 3562
Dd 5049 ggtcttgcacaaagagatggtgcacttgttccctatgagcagaagcgaaaaaccaag 5108
Qy 3563 accaaagtttgacattgacatgacatgaaacaactcgcatatggaacttactgatgggaaagg 3622
Dd 5109 accaaagtttgacattgacatgaaacaactcgcatatggaacttactgatgggaaagg 5168
Qy 3623 agatgaaaaaagagggatgaagagaaggataaaaaagaaagagaagtggtgggaagaga 3682
Dd 5169 agatgaaaaaagagggatgaagagaaggataaaaaagaaagagaagtggtgggaagaga 5228
Qy 3683 aagaagagtctccgaggaagggtgattctctcatcgctcgcatgcacctgggacac- 3739
Dd 5229 aagaagagtctccgaggaagggtgattctctcatcgctcgcatgcacctgggacacag 5288
Qy 3740 ----- 3739
Dd 5289 tgaagatccacttctcttcaactccatctttattcacacaaattagtagaatactcaa 5348
Qy 3740 ----- 3739
Dd 5349 aaatgatgtttgttgcaaaattttaaaaatcactagtttaaccatgtcacaataatttc 5408
Qy 3740 ----- 3739
Dd 5409 ataatgcactctgtgaagaacagggtgacatttatgtgtgacagctgaatggtttatgtgc 5468
Qy 3740 ----- aggagataga 3749
Dd 5469 ctattatttttactgctatagatgaccaaattgaaacttaaacgtttacaggsgagataga 5528
Qy 3750 cgttttcgcacatggaaggatcggtgtgtgattcgcgttcattgagatttcctttacacag 3809
Dd 5529 cgttttcgcacatggaaggatcggtgtgtgattcgcgttcattgagatttcctttacacag 5588
Qy 3810 aatgctcgcgatcaccttttca- 3830
Dd 5589 aatgctcgcgatcacctttcaaggatatagttgcttcaataaaatgagttccaaaaaca 5648
Qy 3831 ----- agctcgtcgttcattgctcttagctgtcgtc 3859
Dd 5649 tagaaaattaacccatggtgtttttacaaatcgactcgcgttcattgctcttagctgtcgtc 5708
Qy 3860 attcctcccaaaattaaagcagcagccgagagaagtaagaaatgttagaagcgtagttgt 3919
Dd 5709 attcctcccaaaattaaagcagcagccgagagaagtaagaaatgttagaagcgtagttgt 5768
Qy 3920 tgaagatcccaagaggatgcatttctgaacttaaatgaaattccttcgtggcaggaaaaaggt 3979
Dd 5769 tgaagatcccaagaggatgcatttctgaacttaaatgaaattccttcgtggcaggaaaaaggt 5828
Qy 3980 tcaacatccatctgcacatggaagtcttctgggttgatagtggaatcaaaaagacagctaaag 4039
Dd 5829 tcaacatccatctgcacatggaagtcttctgggttgatagtggaatcaaaaagacagctaaag 5888
Qy 4040 ggactgttcaaaactctgaaattgaaagatttaatttcttagagagaagagatttcaaaattt 4099
Dd 5889 ggactgttcaaaactctgaaattgaaagatttaatttcttagagagaagagatttcaaaattt 5948
Qy 4100 agaagaggaagtattatcatcaagaattctttgtatcccgcgatatttcagtcggtg 4159
Dd 5949 agaagaggaagtattatcatcaagaattctttgtatcccgcgatatttcagtcggtg 6008
Qy 4160 gagagtggatcctgttcatgtttccaaatcagacgcagagtttctctacaaccagggtgtga 4219
Dd 6009 gagagtggatcctgttcatgtttccaaatcagacgcagagtttctctacaaccagggtgtga 6068
Qy 4220 acaaaaaactgtaagtggaaacatcaaatcagtgcaaaactgggagcccaacttctgtga 4279
Dd 6069 acaaaaaactgtaagtggaaacatcaaatcagtgcaaaactgggagcccaacttctgtga 6128

Qy	4280	tgaatttgccttccaagggaatgagagccgcatactatatgaaggactctggtgatgttca	4339
Db	6129	tgaatttgccttccaagggaatgagagccgcatactatatgaaggactctggtgatgttca	6188
Qy	4340	gaacaagaatactacaataatgtcgctcagaagaataacctgatcttgaataaacaatgaattg	4399
Db	6189	gaacaagaatactacaataatgtcgctcagaagaataacctgatcttgaataaacaatgaattg	6248
Qy	4400	gaagactctgtctgttttggctcagccaagaataatgatactacttggcaacaactccttc	4459
Db	6249	gaagactctgtctgttttggctcagccaagaataatgatactacttggcaacaactccttc	6308
Qy	4460	cagcagctatgagcagtgctgcactcgacgcagccacatgtactacatagaggaattttgg	4519
Db	6309	cagcagctatgagcagtgctgcactcgacgcagccacatgtactacatagaggaattttgg	6368
Qy	4520	aatgcgaagtgaaagccttggttattcttggatgtccatctcacaagagattgaacagatt	4579
Db	6369	aatgcgaagtgaaagccttggttattcttggatgtccatctcacaagagattgaacagatt	6428
Qy	4580	aaagaacaaaaatgtaccacgaggtttttcagacaaggtggaagtggtccaagagaatt	4639
Db	6429	aaagaacaaaaatgtaccacgaggtttttcagacaaggtggaagtggtccaagagaatt	6488
Qy	4640	cacagctcagatcatcaccatcaacgcctcatgaattaccaggaattggattgctcggttc	4699
Db	6489	cacagctcagatcatcaccatcaacgcctcatgaattaccaggaattggattgctcggttc	6548
Qy	4700	ctcaagcccgctccaagaacacacagagagatataccaaacataatacaacaagatgagatgaa	4759
Db	6549	ctcaagcccgctccaagaacacacagagagatataccaaacataatacaacaagatgagatgaa	6608
Qy	4760	taaagatccattttcaaaaaaacatttttgatctgctcaactcctctgaagaatgcct	4819
Db	6609	taaagatcccatctcaaaaaaacatttttgatctgctcaactcctctgaagaatgcct	6668
Qy	4820	tacaagacgtccagtaccaaacagaaacatcacgagtggtctctaccgagagatagaac	4879
Db	6669	tacaagacgtccagtaccaaacagaaacatcacgagtggtctctaccgagagatagaac	6728
Qy	4980	tgctgaagacgtggttgatccgctcagtaacaattccaagcttacagaacatatgtgtcga	4939
Db	6729	tgctgaagacgtggttgatccgctcagtaacaattccaagcttacagaacatatgtgtcga	6788
Qy	4940	atcaaatccagcaataaagagcagacgagcaggttgaaatacaagagacaaaatgccactat	4999
Db	6789	atcaaatccagcaataaagagcagacgagcaggttgaaatacaagagacaaaatgccactat	6848
Qy	5000	tttacgagagatgaaggagcagcttgctgatgggaaaaagcctcaagccagtggtgtag	5059
Db	6849	tttacgagagatgaaggagcagcttgctgatgggaaaaagcctcaagccagtggtgtag	6908
Qy	5060	tctcagaaaaagatgtggaggggaatgaaggagacaggaacgaaaaacaatatggga	5119
Db	6909	tctcagaaaaagatgtggaggggaatgaaggagacaggaacgaaaaacaatatggga	6968
Qy	5120	ttctatagactatgaagcaataaagacgtgctagtatcgccagatcttctgaggtatcaa	5179
Db	6969	ttctatagactatgaagcaataaagacgtgctagtatcgccagatcttctgaggtatcaa	7028
Qy	5180	ggaaaggggatgaataaacatgttgccgctacgaattaaagatttctctagacggatagt	5239
Db	7029	ggaaaggggatgaataaacatgttgccgctacgaattaaagatttctctagacggatagt	7088
Qy	5240	taa 5242	
Db	7089	tga 7091	
RESULT 3			
AAS96686			
ID AAS96686 standard; DNA; 12785 BP.			
XX			

AAS95686;
26-FEB-2002 (first entry)
Arabidopsis DNA for Demeter protein.
Demeter; DMT; Atropos; ATR; 5-methylcytosine glycosylase; ds;
DNA demethylation; transgenic plant; transcription modulation;
flowering time; endosperm development; MEDEA.
Arabidopsis Thaliana.
W0200180626-A1.
01-NOV-2001.
23-APR-2001; 2001WO-US13059.
21-APR-2000; 2000US-0553690.
(REGC) UNIV CALIFORNIA.
Fischer RL, Choi Y, Hannon M, Okamuro JK, Tatarinova TV;
WPI; 2002-055307/07.
P-PSDB; AAU72736.
New polynucleotide that control plant development comprising a sequence
having a specific homology to DEMETER domains A,B or C
Example 1; Page 46-54; 109pp; English.
The invention relates to an isolated polynucleotide sequence or their
complement encoding a polypeptide having a sequence at least 40%
identical to DMT (DEMETER, previously known as ATRPOS (ATR)) Domain A,
or C or their combinations. Also included are an expression
cassette comprising the polynucleotide or comprising a heterologous
polynucleotide under the control of a promoter at least 70% identical to
DMT 5' flanking sequence, DMT 3' flanking sequence or an 5' untranslated
region of DMT, a host cell comprising an exogenous polynucleotide
encoding a DMT-like protein and a transgenic plant comprising a
polynucleotide encoding a DMT-like protein. The expression cassette is
useful for modulating transcription. The method comprises introducing
the cassette into a host cell preferably Agrobacterium by sexual
cross, and selecting a host cell with modulated transcription, where
the protein is capable of exhibiting at least one of the following
biological activities, which include enhanced expression of the
protein in a plant results in a delay in flowering time, introduction of
the protein into a cell results in modulation of methylation of
chromosomal DNA in the cell, reduction of expression of the protein in a
plant results in enhanced endosperm development and expressing of the
protein in an Arabidopsis leaf results in expression of the MEDEA gene.
The polynucleotide is useful for detecting a nucleic acid in a sample.
DEMETER is related to 5-methylcytosine glycosylases and regulates
transcription of target genes by demethylation. The present
invention encodes a DMT protein.
Sequence 12785 BP; 4320 A; 2315 C; 2496 G; 3654 T; 0 other

	Query Match	69.7%;	Score 4788.6;	DB 24;	Length 12785;
	Best Local Similarity	92.6%;	Pred. No. 0;		
	Matches 5228;	Conservative 0;	Mismatches 14;	Indels 401;	Gaps 3;
Qy	1	gtttccggcattgactcgctgagaatcagaaacgttagatcgtgagcttttagctcc	60		
Db	3425	gttccggcattgactcgctgagaatcagaaacgttagatcgtgagcttttagctcc	3484		
Qy	61	atttcctgtttattcacatatattctctttttctctcccccttttttatcttggaatt	120		
Db	3485	atttcctgtttattcacatatattctctttttctctcccccttttttatcttggaatt	3544		
Qy	121	tgtctgtgtaaattttcgcagctgttacattttccgatcacgagaagaatactagggtttt	180		

Db 3545 tttttctgtataattttccagctgttaacattttccgatcacagagaagaataactcctgggtttt 3604
Qy 181 tatgttaatacatactgttctctgtttttctgtatcataataatctcagctattacaacctgat 240
Db 3605 tatgttaatacatactgttctctgtttttctgtatcataaattctcagctattacaacctgat 3664
Qy 241 tttgattctgcgttaataaaacctctgatttgcgtttttattcttcttcaatccccataaaacat 300
Db 3665 tttgattctgcgttaataaaacctctgatttgcgtttttattcttcttcaatccccataaaacat 3724
Qy 301 tgcctactttattcgtctctcttttaccgtttccagcttaaaaaattcttctgctattcaat 360
Db 3725 tgcctactttattcgtctctcttttaccgtttccagcttaaaaaattcttctgctattcaat 3784
Qy 361 gtgtttctcgttttctgtatgagaaaaatactgacaaaaatcatttatttgcattttat 420
Db 3785 gtgtttctcgttttctgtatgagaaaaatactgacaaaaatcatttatttgcattttat 3844
Qy 421 ggtgcagattcttagttaatgtcgccttctctaaaccaagtcagattaaaaaggagtgctc 480
Db 3845 ggtgcagattcttagttaatgtcgccttctctaaaccaagtcagattaaaaaggagtgctc 3904
Qy 481 gtccaatgtctgtttgtttgtgttgagagagattttcggagagatttgcgagtggtgat 540
Db 3905 gtccaatgtctgtttgtgtgttgagagagattttcggagagatttgcgagtggtgat 3964
Qy 541 ttgggggtgagtgatgataaagtttgaagggggagtgattcatcaagtggtttatgaatt 600
Db 3965 ttgggggtgagtgatgataaagtttgaagggggagtgattcatcaagtggtttatgaatt 4024
Qy 601 cgagggtcgatccgggggatagataattttcgaggttctcttgagaatacaactcaacaag 660
Db 4025 cgagggtcgatccgggggatagataattttcgaggttctcttgagaatacaactcaacaag 4084
Qy 661 agttcatgggttcttgggtatccatttacaccaaaaaaccttagatcaagtgctgatggttag 720
Db 4085 agttcatgggttcttgggtatccatttacaccaaaaaaccttagatcaagtgctgatggttag 4144
Qy 721 atgagagagtgataaacaggatctataatgggtttccagggtgggtaattttgacagagg 780
Db 4145 atgagagagtgataaacaggatctataatgggtttccagggtgggtaattttgacagagg 4204
Qy 781 gattctgaacactggttgatgataataatgggggttttttgatcaatggtctcatcaggcg 840
Db 4205 gattctgaacactggttgatgataataatgggggttttttgatcaatggtctcatcaggcg 4264
Qy 841 ttaccaacttaagtatgatcatcaatagcttagcggggtacacatgcacaaagcttggagta 900
Db 4265 ttaccaacttaagtatgatcatcaatagcttagcggggtacacatgcacaaagcttggagta 4324
Qy 901 atagttagagagatcttttgggcaggagtgaggtagcttctctttagcaccagttatca 960
Db 4325 atagttagagagatcttttgggcaggagtgaggtagcttctctttagcaccagttatca 4384
Qy 961 gaaacacacacgggtaatttagagccgggtcaatggaaatttttaactcagatgtgggtatgg 1020
Db 4385 gaaacacacacgggtaatttagagccgggtcaatggaaatttttaactcagatgtgggtatgg 4444
Qy 1021 taaaagggtcttcaaccagagtggcacttctcaagctgggctataatagtgattgaattgg 1080
Db 4445 taaaagggtcttcaaccagagtggcacttctcaagctgggctataatagtgattgaattgg 4504
Qy 1081 atgacttttgaattctctgatcagatgccttctcttcaaaagcttgcgtgagtggtgggg 1140
Db 4505 atgacttttgaattctctgatcagatgccttctcttcaaaagcttgcgtgagtggtgggg 4564
Qy 1141 atagcttattcaaggttcgtcaatgtgagtgatcaaaatctatttcaagtttttttttttc 1200
Db 4565 atagcttattcaaggttcgtcaatgtgagtgatcaaaatctatttcaagtttttttttttc 4624
Qy 1201 cctttctccgtcttgcagtaacttagtagaacaatgaattagaatatcttaagaagt 1260
Db cctttctccgtcttgcagtaacttagtagaacaatgaattagaatatcttaagaagt 5764

Db 4625 cctttctccgtcttgcagtaacttagtagaacaatgaattagaatatcttaagaagaagt 4684
Qy 1261 catggttttgaaacagatggacctccagcgtgtaacaaagcctcttttacaatttgaattcac 1320
Db 4685 catggttttgaaacagatggacctccagcgtgtaacaaagcctcttttacaatttgaattcac 4744
Qy 1321 caattagaagaagacagtcgttgggtcagtcgtgaaagttcgttttcaatatgtaccgtcaa 1380
Db 4745 caattagaagaagacagtcgttgggtcagtcgtgaaagttcgttttcaatatgtaccgtcaa 4804
Qy 1381 cgccagtcgttctcagaacacagtgaaagactggattccttgaacagatagttacaacta 1440
Db 4805 cgccagtcgttctcagaacacagtgaaagactggattccttgaacagatagttacaacta 4864
Qy 1441 ctgacatagaatccacagacgcgaaatctgacaaaagtatgcagagcattatggactcgt 1500
Db 4865 ctgacatagaatccacagacgcgaaatctgacaaaagtatgcagagcattatggactcgt 4924
Qy 1501 ctgctgttaattgcagcggaaagctactgaacaaaatgatggcagcagacagaatgttctcg 1560
Db 4925 ctgctgttaattgcagcggaaagctactgaacaaaatgatggcagcagacagaatgttctcg 4984
Qy 1561 agttcgacttacaacaaactcctcagcagaaacccctcaaaaggaaggaagttcatgc 1620
Db 4985 agttcgacttacaacaaactcctcagcagaaacccctcaaaaggaaggaagttcatgc 5044
Qy 1621 ccaaggtggtcgttggaaagccaaacttaaaagaagccacgcgaaacccctcagaacttccca 1680
Db 5045 ccaaggtggtcgttggaaagccaaacttaaaagaagccacgcgaaacccctcagaacttccca 5104
Qy 1681 aagtggtcgttggaaagccaaacttaaaagaagccacgcgaaagcgtgcaactcaggaaaaag 1740
Db 5105 aagtggtcgttggaaagccaaacttaaaagaagccacgcgaaagcgtgcaactcaggaaaaag 5164
Qy 1741 tgaactcaaaagaaacccgggagtgccaaaagaaaaatttcaagaaatcagcaactaaaa 1800
Db 5165 tgaactcaaaagaaacccgggagtgccaaaagaaaaatttgaagaaatcagcaactaaaa 5224
Qy 1801 agccagccaaatgttggagatagatgagcaacaaagccctgaagtacactcaaaagttgca 1860
Db 5225 agccagccaaatgttggagatagatgagcaacaaagccctgaagtacactcaaaagttgca 5284
Qy 1861 gaaaagcttttgaattttgacttggagaatctcgtgagatgcgaggcaaggtgactctgagt 1920
Db 5285 gaaaagcttttgaattttgacttggagaatctcgtgagatgcgaggcaaggtgactctgagt 5344
Qy 1921 ctgaaatttgcagaaacagtagtggcgcaaaactcgtttctgagatcagagatgccaattg 1980
Db 5345 ctgaaatttgcagaaacagtagtggcgcaaaactcgtttctgagatcagagatgccaattg 5404
Qy 1981 gtgaaactaaatgttggtagtttctcgtgattcagtgtcacaaatagacaagaccaatggattgg 2040
Db 5405 gtgaaactaaatgttggtagtttctcgtgattcagtgtcacaaatagacaagaccaatggattgg 5464
Qy 2041 ggggtatgaacacagcccaacttgaagtgtcaatgggaaacacagccagataaactatctacag 2100
Db 5465 ggggtatgaacacagcccaacttgaagtgtcaatgggaaacacagccagataaactatctacag 5524
Qy 2101 gagcgaaactggccagagacacaaacctgatttattgactagaaaaccagcaatggccagt 2160
Db 5525 gagcgaaactggccagagacacaaacctgatttattgactagaaaaccagcaatggccagt 5584
Qy 2161 tcccagttggcaaccagacaacccagttcccgaatggaaaaaccaaacagcttggcttcaga 2220
Db 5585 tcccagttggcaaccagacaacccagttcccgaatggaaaaaccaaacagcttggcttcaga 5644
Qy 2221 tgaaaaaaccaacttattggctttccatttggtaaccagcaacccctgatgaccataagaa 2280
Db 5645 tgaaaaaaccaacttattggctttccatttggtaaccagcaacccctgatgaccataagaa 5704
Qy 2281 accagcagccttgccttggccatgggttaatacaaaccttatgtatctgataggaacctcac 2340
Db accagcagccttgccttggccatgggttaatacaaaccttatgtatctgataggaacctcac 5764

Db 7925 ||||| agagaggaagtattatcatcacagaattctttgatccggcgatatttcagtcgtg 7984
Qy 4160 gagagttggtacctgttcatttccaaatcagacgacagagtttctctacaaccaggtgtga 4219
Db 7985 gagagttggtacctgttcatttccaaatcagacgacagagtttctctacaaccaggtgtga 8044
Qy 4220 aacaaaactgtcagtggaacatcacatcagtcgaactgggagcccaaaacttgtctga 4279
Db 8045 aacaaaactgtcagtggaacatcacatcagtcgaactgggagcccaaaacttgtctga 8104
Qy 4280 tgaattttgtctcaagggaatgagagacgcacatctatatgaaaggatctgtgtatgtca 4339
Db 8105 tgaattttgtctcaagggaatgagagacgcacatctatatgaaaggatctgtgtatgtca 8164
Qy 4340 gaaacaagaactcaaatgtcgtcgaagaacacctgatcttgaaaaaacaatgaattg 4399
Db 8165 gaaacaagaactcaaatgtcgtcgaagaacacctgatcttgaaaaaacaatgaattg 8224
Qy 4400 gaaagactctgtctgtttgttcgagccaaagaatgatactaatgtggcaacaactcttc 4459
Db 8225 gaaagactctgtctgtttgttcgagccaaagaatgatactaatgtggcaacaactcttc 8284
Qy 4460 cagcagctatgagcagtggtcgactcgacagccacatgtactagacatagaggattttgg 4519
Db 8285 cagcagctatgagcagtggtcgactcgacagccacatgtactagacatagaggattttgg 8344
Qy 4520 aatcgaagtggaagccctgtgtattcttggatgtccatctcacaagagtgacagagt 4579
Db 8345 aatcgaagtggaagccctgtgtattcttggatgtccatctcacaagagtgacagagt 8404
Qy 4580 aaagaacaaaatgtaccacgcaggtttttcagacaaggtggaagtgttccaagagaatt 4639
Db 8405 aaagaacaaaatgtaccacgcaggtttttcagacaaggtggaagtgttccaagagaatt 8464
Qy 4640 cacaggtcagatcattacacatcaacgcctcatggaattccaggaatggattgtccggttc 4699
Db 8465 cacaggtcagatcattacacatcaacgcctcatggaattccaggaatggattgtccggttc 8524
Qy 4700 ctcaagcgctgtccaagaacaccagagcgtatcccaacataatcaacaagatgagatgaa 4759
Db 8525 ctcaagcgctgtccaagaacaccagagcgtatcccaacataatcaacaagatgagatgaa 8584
Qy 4760 taaagcatccatttacaacaaacattttgtgattgtctcaactcctctgaagaatgcct 4819
Db 8585 taaagcatccatttacaacaaacattttgtgattgtctcaactcctctgaagaatgcct 8644
Qy 4820 tacaagcagtcagtaacaaacagacatacaggtatggctgtctacggagagatagaac 4879
Db 8645 tacaagcagtcagtaacaaacagacatacaggtatggctgtctacggagagatagaac 8704
Qy 4880 tgcgtgaagcgtgtgtatccgctcagtaacaaattcaagcttcagaaacatattggtcga 4939
Db 8705 tgcgtgaagcgtgtgtatccgctcagtaacaaattcaagcttcagaaacatattggtcga 8764
Qy 4940 atcaaatccagcaataaagagcagcggcagttgaaatacgaagagacaaatgcactat 4999
Db 8765 atcaaatccagcaataaagagcagcggcagttgaaatacgaagagacaaatgcactat 8824
Qy 5000 tttaagagagatgaaaggacgttgtgtatgggaaagaagcctcacaagcagtggtgatag 5059
Db 8825 tttaagagagatgaaaggacgttgtgtatgggaaagaagcctcacaagcagtggtgatag 8884
Qy 5060 tctcagaaaaagatgtgaggggaatgaaggagagagagaaacgaaacaaacaaataggga 5119
Db 8885 tctcagaaaaagatgtgaggggaatgaaggagagagagaaacgaaacaaacaaataggga 8944
Qy 5120 ttccatagactatgaagcaataagacgtgttagtatcagcgagagattttcagggcgtatcaa 5179
Db 8945 ttccatagactatgaagcaataagacgtgttagtatcagcgagagattttcagggcgtatcaa 9004
Qy 5180 ggaagagggtatgaataacatgttggccgtacgaattaaagagatttcttagaacggtatgt 5239
|||||

Db 9005 ggaagagggatgaataacatgttggcgtacgaattaaggaataactactactatttcagt 9064
Qy 5240 taa 5242
Db 9065 tga 9067
RESULT 4
AAS96690
ID AAS96690 standard; CDNA; 1478 BP.
XX AAS96690;
AC AAS96690;
XX
DT 26-FEB-2002 (first entry)
XX
DE Arabidopsis DMT 5' untranslated region.
XX
KW Demeter; DMT; Atropos; ATR; 5-methylcytosine glycosylase; ss;
KW DNA demethylation; transgenic plant; transcription modulation;
KW flowering time; endosperm development; MEDEA; 5' untranslated region.
XX
OS Arabidopsis Thaliana.
XX
PN W0200180626-Al.
XX
PD 01-NOV-2001.
XX
PF 23-APR-2001; 2001WO-US13059.
XX
PR 21-APR-2000; 2000US-0553690.
XX
PA (REGC) UNIV CALIFORNIA.
XX
PI Fischer RL, Choi Y, Hannon M, Okamura JK, Tatarinova TV;
XX
DR WPI; 2002-055307/07.
XX
PT New polynucleotide that control plant development comprising a sequence
having a specific homology to DEMETER domains A,B or C
PS
XX
XX Claim 11; Page 62; 109pp; English.
XX
CC The invention relates to an isolated polynucleotide sequence or their
complement encoding a polypeptide having a sequence at least 40%
CC identical to DMT (DEMETER, previously known as ATROPOS (ATR)) Domain A, B
CC or C or their combinations. Also included are an expression
CC cassette comprising the polynucleotide or comprising a heterologous
polynucleotide under the control of a promoter at least 70% identical to
CC DMT 5' flanking sequence, DMT 3' flanking sequence or an 5' untranslated
CC region of DMT, a host cell comprising an exogenous polynucleotide
CC encoding a DMT-like protein and a transgenic plant comprising a
polynucleotide encoding a DMT-like protein. The expression cassette is
CC useful for modulating transcription. The method comprises introducing
CC the cassette into a host cell preferably Agrobacterium by sexual
cross, and selecting a host cell with modulated transcription, where
CC the protein is capable of exhibiting at least one of the following
biological activities, which include enhanced expression of the
CC protein in a plant results in a delay in flowering time, introduction of
the protein into a cell results in modulation of methylation of
CC chromosomal DNA in the cell, reduction of expression of the protein in a
plant results in enhanced endosperm development and expressing of the
CC protein in an Arabidopsis leaf results in expression of the MEDEA gene.
CC The polynucleotide is useful for detecting a nucleic acid in a sample.
CC DEMETER is related to 5-methylcytosine glycosylases and regulates
transcription of target genes by demethylation. The present
CC sequence represents the 5' untranslated region of the Arabidopsis
DMT gene.
XX
SQ Sequence 1478 BP; 392 A; 261 C; 322 G; 503 T; 0 other;

Query Match 21.5%; Score 1478; DB 24; Length 1478;
Best Local Similarity 100.0%; Pred. No. 0;

Matches 1478; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy 1 gttctcggcaattgactcgcgcctgagaatcagaagcttagatcggtgagcttttagctcc 60
Db 1 gttctcggcaattgactcgcgcctgagaatcagaagcttagatcggtgagcttttagctcc 60
Qy 61 attttctgtttattacatatattttctctttttttttctctctctctcttttttttttgaatt 120
Db 61 attttctgtttattacatatattttctctttttttttctctctctctcttttttttttgaatt 120
Qy 121 tgtctcgtcaaatttccagctgttacattttccgataccagagagaataactggtgtttt 180
Db 121 tgtctcgtcaaatttccagctgttacattttccgataccagagagaataactggtgtttt 180
Qy 181 tatgttaataacatcattctctgtttcttgatcataaattctcagctattaaacctgat 240
Db 181 tatgttaataacatcattctctgtttcttgatcataaattctcagctattaaacctgat 240
Qy 241 tttgattctcgttaataaaacctctgattgttttattcttcttactttccccaataaacat 300
Db 241 tttgattctcgttaataaaacctctgattgttttattcttcttactttccccaataaacat 300
Qy 301 tgcctactttattcgcctctcttttaccgtttccagctaaaaaattcttcgctattcaat 360
Db 301 tgcctactttattcgcctctcttttaccgtttccagctaaaaaattcttcgctattcaat 360
Qy 361 gtgtttctcgttttcttgatgagaaaaatattctgacaaaaaatcatttattgcaatttat 420
Db 361 gtgtttctcgttttcttgatgagaaaaatattctgacaaaaaatcatttattgcaatttat 420
Qy 421 ggtgcagattcttagttaatgtcgcctctctcttaaccagtcagattaaaaaggagtgctc 480
Db 421 ggtgcagattcttagttaatgtcgcctctctcttaaccagtcagattaaaaaggagtgctc 480
Qy 481 gtccatgttgctttgttttgggtgttgagagagatttttcggagaggttagtgagtgatt 540
Db 481 gtccatgttgctttgttttgggtgttgagagagatttttcggagaggttagtgagtgatt 540
Qy 541 ttggggtgaggtgagtgaaggtttgaagggtgagtgattcatcaagtgtgttatgaatt 600
Db 541 ttggggtgaggtgagtgaaggtttgaagggtgagtgattcatcaagtgtgttatgaatt 600
Qy 601 cgagggtgatccgggggatagatatatttcogaggttcctttggagaaatcaaaactcaacaag 660
Db 601 cgagggtgatccgggggatagatatatttcogaggttcctttggagaaatcaaaactcaacaag 660
Qy 661 agttcaggtgtcttgattccatttaccaccccaaaaacctagatcaagtcctgatgtag 720
Db 661 agttcaggtgtcttgattccatttaccaccccaaaaacctagatcaagtcctgatgtag 720
Qy 721 atgagagagtataaaccagatctaattgggtttccagggtggtgaattttgtagacaggg 780
Db 721 atgagagagtataaaccagatctaattgggtttccagggtggtgaattttgtagacaggg 780
Qy 781 gattctgcaaacactggtgtggatcataatgggggttttttgatcatggtgctcatcagggcg 840
Db 781 gattctgcaaacactggtgtggatcataatgggggttttttgatcatggtgctcatcagggcg 840
Qy 841 ttaccacctaagtatgatcataatagcttagcggatcacatgcacaagccttgagta 900
Db 841 ttaccacctaagtatgatcataatagcttagcggatcacatgcacaagccttgagta 900
Qy 901 atagtggagagatcttttgggcaggagtgaggtgacttctctcttttagcaccagttatca 960
Db 901 atagtggagagatcttttgggcaggagtgaggtgacttctctcttttagcaccagttatca 960
Qy 961 gaaacacacccggttaatgtagagccgttcaattggaatttttacttcagatgtgggtatgg 1020
Db 961 gaaacacacccggttaatgtagagccgttcaattggaatttttacttcagatgtgggtatgg 1020
Qy 1021 taaatgctctttaccacagagtgagcacttctcaagcttgccataaagtgttgaattgg 1080
Db 1021 taaatgctctttaccacagagtgagcacttctcaagcttgccataaagtgttgaattgg 1080
```

```
Qy 1081 atgacttgttgaatcctgatacagatgcccttctccttcacaaagcttctgctgagtggtggg 1140
Db 1081 atgacttgttgaatcctgatacagatgcccttctccttcacaaagcttctgctgagtggtggg 1140
Qy 1141 atagcttattcaagggttcgtaaatgtgagtgatcaaatctatttcoagtttttttttttc 1200
Db 1141 atagcttattcaagggttcgtaaatgtgagtgatcaaatctatttcoagtttttttttttc 1200
Qy 1201 cctttctcgcgttcttcagtaacttagatagaacatgaataatatttaagaagaagt 1260
Db 1201 cctttctcgcgttcttcagtaacttagatagaacatgaataatatttaagaagaagt 1260
Qy 1261 catggttttgaacagatggacctccagcgtgttaacagcctcttcaaatgttgaattcac 1320
Db 1261 catggttttgaacagatggacctccagcgtgttaacagcctcttcaaatgttgaattcac 1320
Qy 1321 caattagaagagagcagttgggtcagctctgtgaagttcgttcaaatgtaccgtcaa 1380
Db 1321 caattagaagagagcagttgggtcagctctgtgaagttcgttcaaatgtaccgtcaa 1380
Qy 1381 cgccagctctgttcagaaacaggtgaaaagactggattccttgaacagatagttacaacta 1440
Db 1381 cgccagctctgttcagaaacaggtgaaaagactggattccttgaacagatagttacaacta 1440
Qy 1441 ctggacatgaaatccccagagccgaaatctgacaaaagt 1478
Db 1441 ctggacatgaaatccccagagccgaaatctgacaaaagt 1478
RESULT 5
AAS96707
ID AAS96707 standard; cDNA; 758 BP.
AC AAS96707;
XX
XX 26-FEB-2002 (first entry)
DE Tomato DMT related cDNA sequence #1.
XX
XX Demeter; DMT; Atropos; ATR; 5-methylcytosine glycosylase; ss;
KW DNA demethylation; transgenic plant; transcription modulation;
KW flowering time; endosperm development; MEDEA.
XX
OS Lycopersicon esculentum.
XX
PN WO200180626-A1.
XX
XX 01-NOV-2001.
XX
XX 23-APR-2001; 2001WO-US13059.
XX
XX 21-APR-2000; 2000US-0553690.
XX
PA (REGC ) UNIV CALIFORNIA.
XX
PI Fischer RL, Choi Y, Hannon M, Okamuro JK, Tatarinova TV;
DR WPI; 2002-055307/07.
DR P-PSDB; AAU72757.
XX
PT New polynucleotide that control plant development comprising a sequence
PS having a specific homology to DEMETER domains A,B or C .
XX
XX Disclosure; Page 92-93; 109pp; English.
XX
XX The invention relates to an isolated polynucleotide sequence or their
CC complement encoding a polypeptide having a sequence at least 40%
CC identical to DMT (DEMETER, previously known as ATROPOS (ATR)) Domain A, B
CC or C or their combinations. Also included are an expression
CC cassette comprising the polynucleotide or comprising a heterologous
CC polynucleotide under the control of a promoter at least 70% identical to
CC DMT 5' flanking sequence, DMT 3' flanking sequence or an 5' untranslated
```


Db 484 aacctgataggagacagtcatttggaaacctccgtaacaagcatattcaaaaggttta 543
 Qy 6531 tcaacggagcagatacacagttctgttttggaaagattcgtatgtgccgtggattcgaa 6590
 Db 544 gcaacaccagaaattcaacagtgcttctggagaggggttctgtgctgagcgtttgaa 603
 Qy 6591 cagaagcaagagcaccgcgtccattaatggcaagggttgcaatttccctgcgagc 6644
 Db 604 aggtcaacgagagcaccgcgtccattaatggccagactgcatttcccgcaagc 657

RESULT 8
 AAS96710
 ID AAS96710 standard; cDNA; 766 BP.
 AC AAS96710;
 XX
 XX
 XX
 XX
 XX 26-FEB-2002 (first entry)
 XX
 XX Cotton DMT related cDNA sequence.
 XX
 XX Demeter; DMT; Atropos; ATR; 5-methylcytosine glycosylase; ss;
 KW DNA demethylation; transgenic plant; transcription modulation;
 KW flowering time; endosperm development; MEDEA.
 XX
 XX Gossypium hirsutum.
 OS
 XX
 XX WO200180626-A1.
 PN
 XX
 XX 01-NOV-2001.
 PD
 XX
 XX 23-APR-2001; 2001WO-US13059.
 XX
 XX
 XX 21-APR-2000; 2000US-0553690.
 XX
 XX (REGC) UNIV CALIFORNIA.
 XX
 XX
 XX Fischer RL, Choi Y, Hannon M, Okamuro JK, Tatarinova TV;
 PI
 XX WPI: 2002-055307/07.
 XX
 XX P-PSDB; AAU72760.
 XX
 XX New polynucleotide that control plant development comprising a sequence
 PT having a specific homology to DEMETER domains A,B or C
 XX
 XX Disclosure; Page 94-95; 109pp; English.
 PS
 XX
 XX The invention relates to an isolated polynucleotide sequence or their
 CC complement encoding a polypeptide having a sequence at least 40%
 CC identical to DMT (DEMETER, previously known as ATROPOS (ATR)) Domain A, B
 CC or C or their combinations. Also included are an expression
 CC cassette comprising the polynucleotide or comprising a heterologous
 CC polynucleotide under the control of a promoter at least 70% identical to
 CC DMT 5' flanking sequence, DMT 3' flanking sequence or an 5' untranslated
 CC region of DMT, a host cell comprising an exogenous polynucleotide
 CC encoding a DMT-like protein and a transgenic plant comprising a
 CC polynucleotide encoding a DMT-like protein. The expression cassette is
 CC useful for modulating transcription. The method comprises introducing
 CC the cassette into a host cell preferably Agrobacterium by sexual
 CC cross, and selecting a host cell with modulated transcription, where
 CC the protein is capable of exhibiting at least one of the following
 CC biological activities, which include enhanced expression of the
 CC protein in a plant results in a delay in flowering time, introduction of
 CC the protein into a cell results in modulation of methylation of
 CC chromosomal DNA in the cell, reduction of expression of the protein in a
 CC plant results in enhanced endosperm development and expressing of the
 CC protein in an Arabidopsis leaf results in expression of the MEDEA gene.
 CC The polynucleotide is useful for detecting a nucleic acid in a sample.
 CC DEMETER is related to 5-methylcytosine glycosylases and regulates
 CC transcription of target genes by demethylation. The present
 CC sequence encodes a DMT-like protein.

SQ Sequence 766 BP; 257 A; 162 C; 166 G; 181 T; 0 other;
 Query Match 4.9%; Score 338.8; DB 24; Length 766;
 Best Local Similarity 70.4%; Pred. No. 1.1e-75;
 Matches 510; Conservative 0; Mismatches 207; Indels 7; Gaps 4;
 Qy 5812 cgaatagagaaactgtgaaccaataattgaagcggcctcgccgggcaagagtga 5871
 Db 34 ctaatgttgcaacaatgaacctatcattgagggcgtgcaaaccttgaaacagagcatg 93
 Qy 5872 ctgaataaacgcgagtgatattgaagatgcttactacaattgaggaacctgacgagatcc 5931
 Db 94 cagaaggtacagagtgatattgaagatgcaagct---atgatgccagatgaaattc 150
 Qy 5932 caacaataaaactcaacattgaacagtttggaatgactctctacggaacacatgaaagaa 5991
 Db 151 ccaataaaaaactcaacattgaagagttcacagcaaacctcacgattacattcagcaggca 210
 Qy 5992 acatggagctccaagaaggtgacatgtcccaaggctttgggtgttgcataccaacaacta 6051
 Db 211 atatggaactccaagaaggggactgttcaaaagcttttagtagctttgaatcctgatgctg 270
 Qy 6052 cttctattccaactcccaactaaagaacattagcgtctctcagacagagacccaagtgt 6111
 Db 271 cttctatccctactccaaaattggaagaatgtaagcaggctcagcaacagagcactatgat 330
 Qy 6112 acgagctccagattcacatcgtctccttgatggtatggataaaagagaaacagatgac 6171
 Db 331 atgagcttccagataaacatcctctcttgaacagagtggaagagcgggaacctgatgac 390
 Qy 6172 caagctcttatctcttagctatatggacaccaggtgaaacagcgaattcggcacaaccgc 6231
 Db 391 ctgacctctatcttcttgcaataatggacaccaggtgaaactgcaaaactcaattcaaccac 450
 Qy 6232 ctgaacagaagtgtagggggaaagcgtctctgcaaaatgtgctttgacgagagactgttctg 6291
 Db 451 cagaacaaagtgtgggtcccaagaaccaggagagactgtgcaatgagagagacctgtttg 510
 Qy 6292 agtgaacagctcaggggaagcaaacctcacagacagcttcgaggaactctctgatacctt 6351
 Db 511 ctgcaacagctgtaagagagagactaacacactgagacagtcggaggaaacctctgatacctt 570
 Qy 6352 gtcggactgccatgagaggaagttttccgctcaacgggacata---tttccaagtcaacga 6409
 Db 571 gtagaatgcaatgagaggaagcttttcccttaattggaacttaatttcaagttaagag 630
 Qy 6410 gttattgacagaccagtgctccagtcctcaaacccatcgatgttcccta-gagattggatat 6468
 Db 631 gtctttgcatgatcatgaatcaagcctcaacccaatgacgttccaagggaatggattg 690
 Qy 6469 gggatctcccaagaa-ggactgttttacttcggaaacatcagtaacatcaatttcagaggt 6527
 Db 691 ggaattaaacaagaacgaactgtatacttggaaacatcctcgtggttcataatttaaggga 750
 Qy 6528 cttt 6531
 Db 751 cttt 754

RESULT 9
 AAS96703
 ID AAS96703 standard; cDNA; 1134 BP.
 XX
 XX AAS96703;
 AC
 XX
 XX 26-FEB-2002 (first entry)
 XX
 XX Soybean DMT2 DNA.
 XX
 XX Demeter; DMT; Atropos; ATR; 5-methylcytosine glycosylase; ss;
 KW DNA demethylation; transgenic plant; transcription modulation;
 KW flowering time; endosperm development; MEDEA.

XX OS Glycine max.
 XX PN WO200180626-A1.
 XX XX
 XX PD 01-NOV-2001.
 XX PF 23-APR-2001; 2001WO-US13059.
 XX PR 21-APR-2000; 2000US-0553690.
 XX PA (REGC) UNIV CALIFORNIA.
 XX PI Fischer RL, Choi Y, Hannon M, Okamuro JK, Tatarinova TV;
 XX PT WPI; 2002-055307/07.
 XX DR P-PSDB; AAU72753.
 XX PT
 XX PT New polynucleotide that control plant development comprising a sequence
 XX PT having a specific homology to DEMETER domains A,B or C
 XX PS Disclosure; Page 89-90; 109pp; English.
 XX CC The invention relates to an isolated polynucleotide sequence or their
 XX CC complement encoding a polypeptide having a sequence at least 40%
 XX CC identical to DMT (DEMETER, previously known as ATROPOS (ATR)) Domain A, B
 XX CC or C or their combinations. Also included are an expression
 XX CC cassette comprising the polynucleotide or comprising a heterologous
 XX CC polynucleotide under the control of a promoter at least 70% identical to
 XX CC DMT 5' flanking sequence, DMT 3' flanking sequence or an 5' untranslated
 XX CC region of DMT, a host cell comprising an exogenous polynucleotide
 XX CC encoding a DMT-like protein and a transgenic plant comprising a
 XX CC polynucleotide encoding a DMT-like protein. The expression cassette is
 XX CC useful for modulating transcription. The method comprises introducing
 XX CC the cassette into a host cell preferably Agrobacterium by sexual
 XX CC cross, and selecting a host cell with modulated transcription, where
 XX CC the protein is capable of exhibiting at least one of the following
 XX CC biological activities, which include enhanced expression of the
 XX CC protein in a plant results in a delay in flowering time, introduction of
 XX CC the protein into a cell results in modulation of methylation of
 XX CC chromosomal DNA in the cell, reduction of expression of the protein in a
 XX CC plant results in enhanced endosperm development and expressing of the
 XX CC protein in an Arabidopsis leaf results in expression of the MEDEA gene.
 XX CC The polynucleotide is useful for detecting a nucleic acid in a sample.
 XX CC DEMETER is related to 5-methylcytosine glycosylases and regulates
 XX CC transcription of target genes by demethylation. The present
 XX CC sequence encodes a DMT-like protein.
 XX SQ Sequence 1134 BP; 384 A; 216 C; 264 G; 270 T; 0 other;

Query Match 4.8%; Score 329; DB 24; Length 1134;
 Best Local Similarity 70.0%; Pred. No. 4.1e-73;
 Matches 443; Conservative 0; Mismatches 190; Indels 0; Gaps 0;

Qy 4973 tgaatacaagagagacaatgccactattttacagagagatgaaggagcgttgctgatgg 5032
 |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
 Db 502 tgaagcagtaagatagattcttccctgtataaataaggagcaggagcatgaaaga 561
 |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
 Qy 5033 gaaaaagcctacagcagtgaggtagtctcagaaaaagatgtagggggaatgaaggag 5092
 |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
 Db 562 gaaaaaataaactttgattgggtagtttaagatacaagcagaagcgaaggcaggaa 621
 |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
 Qy 5093 acaggaacgaacaaacaaatattgattccatagactatgaagcaataagacgtctag 5152
 |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
 Db 622 aagggaaagacagagaacaccatggactcttgactcgtggtatgctgttagacgcgaga 681
 |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
 Qy 5153 tatcagcagatttctgaggctatcaaggaaaggaggaatgaataacatgttggccgtacg 5212
 |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
 Db 682 tgtcagtgaattccaatgcaatcaaaagaaaggcgcatgaacacatgcttgcgtgaacg 741
 |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
 Qy 5213 aattaagattctcagaacggatagtttaagatcatcgttggtatcgaccttgatggtt 5272
 |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||

Db 742 tattcagagtttctcgaatctattggttgacaagcatgggggcatcgtatttgatggtcct 801
 Qy 5273 gagagaatctcctcctgataaagccaaggactatctcttgagcataagaggtctcgggttt 5332
 |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
 Db 802 gagagatgttccacctgatcaagcaaaagaattcttctcagcataagggttggtgatt 861
 Qy 5333 gaaaagtgttgaaatgcgtgcgactcttaacacatcccaatcttcttccctgttgacac 5392
 |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
 Db 862 gaaaagtgtggtgtgtacgactcttaacacatcccaatcttcttccgttcgggtggacac 921
 Qy 5393 gaattgtggaagatagcagattaggtatgggtgcctctacaaacccctacacctgaatc 5452
 |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
 Db 922 aaattgtggacgtatagcagtaagattgggtggtgcctctccagcactgcccagagtc 981
 Qy 5453 acttcagttacacctcctgcagctatataccagtgctgcagtcctccatcaaaaattcttgg 5512
 |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
 Db 982 actacagttgcattcttcttagaattgtaccagtggttgaggatccatacaaaaatactctg 1041
 Qy 5513 gccaaagatttgcaaaactcgatcaacgaacacactgtatgaattacactaccactgattac 5572
 |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
 Db 1042 gccccgctctgcaagctagaccacaaagacatgtatgagtgctaccagctgattac 1101
 Qy 5573 gtttgaaaagttatttgacaaagagtagacc 5605
 |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
 Db 1102 atttggaaggtcttctgtactaaaagcacc 1134
 |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
 RESULT 10
 AAS96704
 ID AAS96704 standard; cDNA; 616 BP.
 XX AC AAS96704;
 XX DT 26-FEB-2002 (first entry)
 XX DE Soybean DMT3 DNA.
 XX KW Demeter; DMT; Atropos; ATR; 5-methylcytosine glycosylase; ss;
 KW DNA demethylation; transgenic plant; transcription modulation;
 KW flowering time; endosperm development; MEDEA.
 XX OS Glycine max.
 XX PN WO200180626-A1.
 XX PD 01-NOV-2001.
 XX PF 23-APR-2001; 2001WO-US13059.
 XX PR 21-APR-2000; 2000US-0553690.
 XX PA (REGC) UNIV CALIFORNIA.
 XX PI Fischer RL, Choi Y, Hannon M, Okamuro JK, Tatarinova TV;
 XX PT WPI; 2002-055307/07.
 XX DR P-PSDB; AAU72754.
 XX PT New polynucleotide that control plant development comprising a sequence
 XX PT having a specific homology to DEMETER domains A,B or C
 XX PS Disclosure; Page 90-91; 109pp; English.
 XX CC The invention relates to an isolated polynucleotide sequence or their
 XX CC complement encoding a polypeptide having a sequence at least 40%
 XX CC identical to DMT (DEMETER, previously known as ATROPOS (ATR)) Domain A, B
 XX CC or C or their combinations. Also included are an expression
 XX CC cassette comprising the polynucleotide or comprising a heterologous
 XX CC polynucleotide under the control of a promoter at least 70% identical to
 XX CC DMT 5' flanking sequence, DMT 3' flanking sequence or an 5' untranslated
 XX CC region of DMT, a host cell comprising an exogenous polynucleotide
 XX CC encoding a DMT-like protein and a transgenic plant comprising a
 XX CC polynucleotide encoding a DMT-like protein. The expression cassette is

CC	protein in a plant results in a delay in flowering time, introduction of			
CC	the protein into a cell results in modulation of expression of the protein in a			
CC	chromosomal DNA in the cell, reduction of expression and expression of the			
CC	plant results in enhanced endosperm development and expression of the			
CC	protein in an Arabidopsis leaf results in expression of the MEDEA gene.			
CC	The polynucleotide is useful for detecting a nucleic acid in a sample.			
CC	DEMETER is related to 5-methylcytosine glycosylases and regulates			
CC	transcription of target genes by demethylation. The present			
CC	sequence encodes a DMT-like protein.			
XX	Sequence 517 BP; 150 A; 110 C; 125 G; 132 T; 0 other;			
SQ				
Query Match				
Best Local Similarity 3.3%; Score 228.2; DB 24; Length 517;				
Matches 343; Conservative 66.3%; Pred. No. 1.1e-47;				
Mismatches 0; Mismatches 173; Indels 1; Gaps 1;				
QY	6111	tacgagctccacagattcaca-tcgtctctcttgatgtgataaaagagaacacgatga	6169	
DB	1	tatgaactgccagattcacacgcctctcttggaagattcgaaacagagagaacacgatga	60	
QY	6170	tccaagtcccttattcttagctatcgacacacaggtgaaacagcgaaattcgcgcacaacc	6229	
DB	61	tcctctgccatattctctttccatattggacccaggtgaaactgcacaatcgatcgatgc	120	
QY	6230	gcctgaacagagagtgaggagaaagcgctctgcgcaaaatgtctttgaacgactgttc	6289	
DB	121	ccccagacattctctggtattccaggggagacggttagactatgtggaagtccaacatgctt	180	
QY	6290	tgaggttaacagctctgagggaagacaaactcacagacagcttcgagggaactcttctgatacc	6349	
DB	181	tagttgcaacaataatacagagaaatgcaggctcagaagaatcagaggaacacttttgatacc	240	
QY	6350	ttctcgactgcctatgagagagaggtttctcgctcaacgggacatatattccaagtcaacga	6409	
DB	241	atgccgaacagcaatgagagagaggtctccacacttaatgggacgtattttcaagtaatga	300	
QY	6410	gttatttgcagaccacagctcagctcacaacccatcgatgttctctagagattgatatg	6469	
DB	301	ggatttgcgtgacctattgctcaagtcacaaatccaattgattgctccacgaagtgtgattg	360	
QY	6470	ggatctcccaagaaggagctgtttacttcggaacatcagtaacatcaatattcagaaggtct	6529	
DB	361	ggacctcccaagacgaaactgtttactttggaacctcagttctctacaataattcagaagttt	420	
QY	6530	tcaacggagcagatatacagttctcttttggaaaggattcgatgtgtccgtggatcga	6589	
DB	421	aacgactgaagagagatacaacagatgttttggagaggtttgttgcgtgaggggctttga	480	
QY	6590	acagaagacaagagcagcgcgtccatttaattggcaagg	6626	
DB	481	taggacagtcgaggcaccacgaaggccctttatgcaagg	517	
RESULT 13				
ID	AAS96711 standard; cDNA; 583 BP.			
XX	AAS96711;			
AC	AC			
XX	26-FEB-2002 (first entry)			
DT	Soybean DMT related cDNA sequence.			
DE	Demeter; DMT; Atropos; ATR; 5-methylcytosine glycosylase; ss;			
KW	DNA demethylation; transgenic plant; transcription modulation;			
KW	flowering time; endosperm development; MEDEA.			
XX	Glycine max.			
OS	WO200180626-A1.			
XX	01-NOV-2001.			
PD				

Search completed: September 27, 2002, 10:16:06
Job time: 15850 sec

Result No.	Score	Query		Length	DB	ID	Description
		Match	%				
1	538	7.8	538	9	AV554948	AV554948 RV554948	
2	533	7.8	533	9	AV554762	AV554762 AV554762	
C 3	528.4	7.7	623	9	AV545918	AV545918 AV545918	
	525.4	7.6	527	10	B65668	B65668 AV556668	
C 5	493.6	7.2	500	12	B60854	B60854 T1907FF TAM	
6	478.6	7.0	486	12	B28303	B28303 T9J15TRB TA	
C 7	477	6.9	575	9	AV545791	AV545791 AV545791	
8	410	6.0	759	10	BG133849	BG133849 EST469495	
9	403.4	5.9	822	12	BH464280	BH464280 BOGY248TF	
10	351.6	5.1	657	9	AW267923	AW267923 EST306265	
11	351.2	5.1	654	10	BG594366	BG594366 EST493044	
12	326.4	4.7	768	10	BF275260	BF275260 GA_Eb002	
13	321.6	4.7	923	9	BE035970	BE035970 MO22F11 M	
14	319.6	4.7	602	9	AW737153	AW737153 EST338580	
15	318.6	4.6	684	10	BG450287	BG450287 NF015A07D	
16	312.2	4.5	537	9	AW092297	AW092297 EST285477	
17	300	4.4	535	10	BF053858	BF053858 EST439004	

```
Best Local Similarity 100.0%; Pred. No. 5.5e-78;
Matches 538; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3088 ctgaggttagaatacagatgacccaactgacgagcaagaaaggttaaaaaactg 3147
Db 1 CTGAGGTTATAGAAATCGAGGATGATCCAACTGATGGGCAAGAAAGGTAAATAACTG 60
QY 3148 ccagcatcagtaaggtgcatctaaagaaactcgtctccagtttaaagacagcagaaa 3207
Db 61 CCAGCATCAGTAAGGTGCATCTAAAGGAACTCGTCTCCAGTTAAAGACACAGCAGAAA 120
QY 3208 aggagaataatgttgtcccaaaacgcctgcacaaaaggggtcgagcaggttagaaaaaat 3267
Db 121 AGGAGAAATGTATTGTCCCAAAACGCTGCAAAAAGGGTCGAGCAGGTAGAAAAAAT 180
QY 3268 cagtacctcgcctgctcatgcctcagagatccagcttttgcaacctactcctccaaaga 3327
Db 181 CAGTACCTCCGCTGCTCATGCCCTCAGAGATCCAGCTTTTGCAACCTACTCCTCCAAAGA 240
QY 3328 cacccttatcaagaagcaagcctaaagaaaggaagagaaagtcctatacaagattcaggaa 3387
Db 241 CACCTTTATCAAGAAGCAAGCCTTAAGGAAGGAGAAAGTCCATACAGATTCAGGAA 300
QY 3388 aagcaagaggttccatcaggagaactctgtgtcaggattctattgcggaaataattaca 3447
Db 301 AAGCAAGAGGTCCATCAGGAGAACTTCTGTCTCAGGATTCTATTGCGGAAATAATTACA 360
QY 3448 ggatgcaaaatctgtatctaggagacaaaagaaagagacaagcaaaatgcaatgtct 3507
Db 361 GGATGCAAAATCTGTATCTAGGAGACAAAGAAAGAGAAAGCAAAATGCAATGTCT 420
QY 3508 tgtacaagagagatggtgcacttttccctatgagagcaagcaagcaagacccca 3567
Db 421 TGTACAAGAGAGATGGTGCACTTGTCTTATGAGAGCAAGAGCAAGCAAGACCCA 480
QY 3568 aagtgacattgacgtgaacaaactcgcatatggaacttactgattgggaaa 3620
Db 481 AAGTTGACATTGAGATGAAACAACACTCGCATATGGAACCTTACTGATGGGAAA 538

RESULT 2
AV554762 AV554762 533 bp mRNA linear EST 06-SEP-2000
LOCUS AV554762 Arabidopsis thaliana roots Columbia Arabidopsis thaliana
DEFINITION cDNA clone RZL01d12R 5', mRNA sequence.
ACCESSION AV554762
VERSION AV554762.1 GI:8726176
KEYWORDS EST.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 533)
AUTHORS Asamizu,E., Nakamura,Y., Sato,S. and Tabata,S.
TITLE A large scale analysis of cDNA in Arabidopsis thaliana: Generation
of 12,028 non-redundant expressed sequence tags from normalized and
size-selected cDNA libraries
JOURNAL DNA Res. 7, 175-180 (2000)
MEDLINE 20363093
COMMENT Contact: Erika Asamizu
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: asamizuekazusa.or.jp, URL:http://www.kazusa.or.jp/en/plant/.
Location/Qualifiers
1. 533
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/clone="RZL01d12R"
/clone_lib="Arabidopsis thaliana roots Columbia"
/tissue_type="roots"

FEATURES
source
Location/Qualifiers
1. 533
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/clone="RZL01d12R"
/clone_lib="Arabidopsis thaliana roots Columbia"
/tissue_type="roots"

Query Match 7.8%; Score 533; DB 9; Length 533;
Best Local Similarity 100.0%; Pred. No. 3.6e-77;
Matches 533; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3088 ctgaggttagaatacagatgacccaactgacgagcaagaaaggttaaaaaactg 3147
Db 1 CTGAGGTTATAGAAATCGAGGATGATCCAACTGATGGGCAAGAAAGGTAAATAACTG 60
QY 3148 ccagcatcagtaaggtgcatctaaagaaactcgtctccagtttaaagacagcagaaa 3207
Db 61 CCAGCATCAGTAAGGTGCATCTAAAGGAACTCGTCTCCAGTTAAAGACACAGCAGAAA 120
QY 3208 aggagaataatgttgtcccaaaacgcctgcacaaaaggggtcgagcaggttagaaaaaat 3267
Db 121 AGGAGAAATGTATTGTCCCAAAACGCTGCAAAAAGGGTCGAGCAGGTAGAAAAAAT 180
QY 3268 cagtacctcgcctgctcatgcctcagagatccagcttttgcaacctactcctccaaaga 3327
Db 181 CAGTACCTCCGCTGCTCATGCCCTCAGAGATCCAGCTTTTGCAACCTACTCCTCCAAAGA 240
QY 3328 cacccttatcaagaagcaagcctaaagaaaggaagagaaagtcctatacaagattcaggaa 3387
Db 241 CACCTTTATCAAGAAGCAAGCCTTAAGGAAGGAGAAAGTCCATACAGATTCAGGAA 300
QY 3388 aagcaagaggttccatcaggagaactctgtgtcaggattctattgcggaaataattaca 3447
Db 301 AAGCAAGAGGTCCATCAGGAGAACTTCTGTCTCAGGATTCTATTGCGGAAATAATTACA 360
QY 3448 ggatgcaaaatctgtatctaggagacaaaagaaagagacaagcaaaatgcaatgtct 3507
Db 361 GGATGCAAAATCTGTATCTAGGAGACAAAGAAAGAGAAAGCAAAATGCAATGTCT 420
QY 3508 tgtacaagagagatggtgcacttttccctatgagagcaagcaagcaagacccca 3567
Db 421 TGTACAAGAGAGATGGTGCACTTGTCTTATGAGAGCAAGAGCAAGCAAGACCCA 480
QY 3568 aagtgacattgacgtgaacaaactcgcatatggaacttactgattgggaaa 3620
Db 481 AAGTTGACATTGAGATGAAACAACACTCGCATATGGAACCTTACTGATGGGAAA 538

RESULT 3
AV545918/c AV545918 623 bp mRNA linear EST 06-SEP-2000
LOCUS AV545918 Arabidopsis thaliana roots Columbia Arabidopsis thaliana
DEFINITION cDNA clone RZL05C12F 3', mRNA sequence.
ACCESSION AV545918
VERSION AV545918.1 GI:8717332
KEYWORDS EST.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 623)
AUTHORS Asamizu,E., Nakamura,Y., Sato,S. and Tabata,S.
TITLE A large scale analysis of cDNA in Arabidopsis thaliana: Generation
of 12,028 non-redundant expressed sequence tags from normalized and
size-selected cDNA libraries
JOURNAL DNA Res. 7, 175-180 (2000)
MEDLINE 20363093
COMMENT Contact: Erika Asamizu
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: asamizuekazusa.or.jp, URL:http://www.kazusa.or.jp/en/plant/.
Location/Qualifiers
```


[illegible]


```

Db 301 TTTAAACAGATAGTTACAACTACTGGACATGAATCCAGCGCGAGATCTGACAAAGTA 360
QY 1480 tgcagagcattatggactcgtcgtgtaatgcagcgaagctactgaacaaatgatg 1539
Db 361 TGCAGAGCATTATGGACTCGTCTGTTTAATGCCAGCGGAAGCTACTGAACAAATGATG 420
QY 1540 gcagcagacaagatgtctcgtgagtgacaccttaacaaaactcctcagcagaacccctcca 1599
Db 421 GCAGCAGACAAGATGTTCTGGAGTTGGACCTTAACAAAACTCCTCTGCAGAACCCCTCCA 480
QY 1600 aaagga 1605
Db 481 NAAGGA 486

RESULT 7
LOCUS AV545791/c 575 bp mRNA linear EST 06-SEP-2000
DEFINITION AV545791 Arabidopsis thaliana roots Columbia Arabidopsis thaliana
cDNA clone RZL01d12F 3', mRNA sequence.
ACCESSION AV545791
VERSION AV545791.1 GI:8717205
KEYWORDS EST
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 575)
REFERENCE
AUTHORS Asanizu,E., Nakamura,Y., Sato,S. and Tabata,S.
TITLE A large scale analysis of cDNA in Arabidopsis thaliana: Generation
of 12,028 non-redundant expressed sequence tags from normalized and
size-selected cDNA libraries
JOURNAL DNA Res. 7, 175-180 (2000)
MEDLINE 20363093
COMMENT Contact: Erika Asanizu
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: asanizu@kazusa.or.jp, URL:http://www.kazusa.or.jp/en/plant/.

FEATURES
source
location/Qualifiers
1..575
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/clone="RZL01d12F"
/clone_lib="Arabidopsis thaliana roots Columbia"
/tissue_type="roots"
/note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
XhoI"
BASE COUNT 167 a 116 c 119 g 173 t
ORIGIN

Query Match 6.9%; Score 477; DB 9; Length 575;
Best Local Similarity 99.8%; Pred. No. 4.7e-68;
Matches 488; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 6365 gagaggaagtcttcgctcaacgggacatatattccaagtccaacgagttatttgcagacca 6424
Db 575 GAGAGGAAGTTTTCGCTCAACGGGACATATTTCCAAAGTCAACGAGTTATTTCGACACCA 516
QY 6425 cgagtcagctcacaacccatcgatgtctcgtagatgtgatatggatctctcccaagaag 6484
Db 515 CGAGTCCAGCTCAAAACCCATCGATGTTCTTAGAGATTGGATGGATCTCCCAAGAAG 456
QY 6485 gactgttacttcggaacatcagtaacataatcagaggtcttccaacggagcagat 6544
Db 455 GACTGTTTACTTCGGAACATCAGTAACATCAATATTCAGAGGTCITTCACGGAGCAGAT 396
QY 6545 acagttctcgttttggaaaggattcgatgtgtcctcgttgattcgacagaagacagagc 6604

```

```

Db 395 ACAGTTCCTCTTTTGGAAAGGATTGCTATGTGTCCGTGGATTGGAACAGAAAGAGAC 336
QY 6605 acgcgctcattaaatggcgaaggttgcatcttctcgcgagcaaatgaagaacaaac 6664
Db 335 ACCGGCTCCATTAAATGGCAAGGTTGCATTTTCTCGAGCAAAATGGAACAACAACAAAC 276
QY 6665 ctaaagatgactgggaagaagaacgcattgctctctcgtctcctcctctatttaagcca 6724
Db 275 CTAAGAATGACTGGAAAGAAAGCAACGCATTGCTTCTCTGCTCTCCTCTATTATTAAGCCA 216
QY 6725 ggaagaagtcctcatttagacataataacgaagaatccaaataggctatttctctcttcttc 6784
Db 215 GGAAAAGTCCCATTTAGACATAATAACAGGAATCCAAATAGGCTATTTTCTCTCTTTTC 156
QY 6785 ttatttcattcatagcagcagcgacac-aaaaagttttttgggtatttattttttctc 6843
Db 155 TTATTTTCATTATAGAGCAGACGACACAAAAAAGTTTTTTGGGTTATTATTTCCTC 96
QY 6844 tctaacaaa 6852
Db 95 TCTAACAAA 87

RESULT 8
LOCUS BG123849
DEFINITION EST469495 tomato shoot/meristem Lycopersicon esculentum cDNA clone
CTOP3G24 5' sequence, mRNA sequence.
ACCESSION BG123849
VERSION BG123849.1 GI:12624037
KEYWORDS EST
SOURCE tomato.
ORGANISM Lycopersicon esculentum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
Lycopersicon.
1 (bases 1 to 759)
REFERENCE
AUTHORS van der Hoeven,R., Bezzerides,J., Sun,H., Cho,J., Utterback,T.,
Hansen,C., Ronning,C. and Tanksley,S.
TITLE Generation of ESTs from tomato shoot/meristem tissue
JOURNAL Unpublished (2001)
COMMENT Contact: CUGI
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Email: http://www.genome.clemson.edu/orders/index.html.

FEATURES
source
location/Qualifiers
1..759
/organism="Lycopersicon esculentum"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone="ctOP3G24"
/clone_lib="tomato shoot/meristem"
/tissue_type="shoot/meristem"
/dev_stage="developing shoots from 4-6wks old plants"
/lab_host="SOLR"
/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
XhoI; Small expanding leaves from the growing tip were
taken from greenhouse plants (4-6wks old TA496). Tissue
was immediately frozen in liquid nitrogen."
BASE COUNT 225 a 158 c 191 g 185 t
ORIGIN

Query Match 6.0%; Score 410; DB 10; Length 759;
Best Local Similarity 73.4%; Pred. No. 3.5e-57;
Matches 524; Conservative 0; Mismatches 190; Indels 0; Gaps 0;

QY 5062 tcagaaaaagatgtggagggaatgaaggagagagcaggaagcaacaaaaacaatatagatt 5121
Db 4 TGAGAAAGGAAGTCCCAATCAAGAGTGGGAAAAAGAAAGCAAGAGTGAATGGACT 63

```



```
QY 6526 gtctttt 6531
| | | | |
Db 751 GACTTT 756

RESULT 13
BE035970
LOCUS MO22F11 MO Mesembryanthemum crystallinum cDNA 5' similar to protein
DEFINITION , mRNA sequence.
ACCESSION BE035970
VERSION BE035970.1 GI:8330979
KEYWORDS EST.
SOURCE common ice plant.
ORGANISM Mesembryanthemum crystallinum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Caryophyllidae; Caryophyllales; Alzooceae; Mesembryanthemum.
REFERENCE 1 (bases 1 to 923)
AUTHORS Bohnert,H.J., Borchert,C., Brazille,S., Brooks,J., Eaton,M., Ferrea
,H., Kwasaki,S., McCollough,A., Michalowski,C.B., Palacio,C.,
Scara,G., Wheeler,M. and Zepeda,G.R.
TITLE Functional Genomics of Plant Stress Tolerance
JOURNAL Unpublished (2000)
COMMENT Contact: Michalowski,C.B.
University of Arizona
Bio Sciences West room 513, Tucson, AZ 85721, USA
Tel: 520-621-7982
Fax: 520-621-1697
Email: cbm@u.arizona.edu
An open reading frame exists.
Location/Qualifiers
1..923
/organism="Mesembryanthemum crystallinum"
/db_xref="taxon:3544"
/clone_lib="MO"
/tissue_type="apical meristem and leaf primordia"
/dev_stage="5 weeks"
/note="no stress"

BASE COUNT 290 a 216 c 202 g 214 t 1 others
ORIGIN

Query Match 4.7%; Score 321.6; DB 9; Length 923;
Best Local Similarity 65.6%; Pred. No. 8.4e-43;
Matches 520; Conservative 0; Mismatches 260; Indels 13; Gaps 3;

QY 5816 tagaataacttgacacaaataattgaagagccgctcgccggggaagagtgcaactga 5875
| | | | |
Db 59 TGGAAATTACTTGTGAGCCAAATTTGTGAGTCCGGCATCACCAGAACCCAGAG--CAGGA 115
| | | | |

QY 5876 aataacgagagtgatattgaagatgcttaactacaatgagacccctgacgagatcccaac 5935
| | | | |
Db 116 ACAACATATTCCGACATTGAAGAT-----TATGAAGATCTGTGATGAATTCACAC 166
| | | | |

QY 5936 aataaactcaacattgaacagtttggaatgactctacgggaacacatggaagaaacat 5995
| | | | |
Db 167 AATCAAGCTCAATATGAAGAATTCACACACTAGAGAAATTTATGCAAGGAATAG 226
| | | | |

QY 5996 ggagtcacaagaagtgacatgtccaaagcttggtgtgcttgcattccaaactacttc 6055
| | | | |
Db 227 AGAACTTCAAGAGACTGCACATATCAAGGGCCCTTGGTGGCTTTAACACCTGAAAGCTGCTC 286
| | | | |

QY 6056 tattccaactcccaactaaagaacattagcgtctcaggacgagacccaagtacga 6115
| | | | |
Db 287 CATTCTACACAAAGCTCAAGAATATCAGTCGGCTAGAACAGAGCATCATGTCTACGA 346
| | | | |

QY 6116 gtccacagattcacatgctctccttgatggtatggaataaaagacacagatgcccaag 6175
| | | | |
Db 347 ACTTCCGACTCACACCCTCTTTTAAAGGATTGGACAAACGAGAGCCGTGATGATCCGTG 406
| | | | |

QY 6176 tccctatctcttagtatatggaacacccaggtgaaacagcgaattcggcacaccocctga 6235
| | | | |
```

```
Db 407 TTCATACCTACTTGCATCTGGACCCCGAGACCCGAAATTCCTCCAACTCCTGA 466
QY 6236 acagaagtgtgagggaaagcgtctgcaaaatgtgctttgacagactttcttgagt 6295
| | | | |
Db 467 AAGAAGATGCAGCTCTGAAGACCCAAACTTACTTTGCAATGAGCAACATGCGCTACTG 526
| | | | |
QY 6296 taacagctgagggagaaacacacagacagcttcgagaaactctctgatacccttgcg 6355
| | | | |
Db 527 TAGCAGTCAGCGGAAGTAACCTCCCAAAATTTGTGGGNGAACTCTTCTTGATACCATGCG 586
| | | | |
QY 6356 gactgccaatgagagaaagtcttcgctcaacgggacacatatcttcaagtcacagagtatt 6415
| | | | |
Db 587 AACAGCAATGAGAGCGAGCTTTCCACATCAATGGAACATACTTTCAAGTAAACGAGGTGT 646
| | | | |
QY 6416 tgcagaccacagtcacagctctcacaacccatcgatgttcctagagattgatatggatct 6475
| | | | |
Db 647 TGCAGACCATGACTCCACCCCTTACCCAGTGGAGTGTGCCAGATCTTTGTGTGGAACCT 706
| | | | |
QY 6476 cccaagaaggactgttacttcggaacatcgataacatcaatttcagagtgcttcttcaac 6535
| | | | |
Db 707 GCCAAGACGAACCTGTATA-TTTTGTACTTCTATATCCACCATATTTCAGAGCTCTGAACAC 765
| | | | |
QY 6536 ggagcagatacagttctcttggaaaggattcgatgtcgcgtgagtcgcgaacagaa 6595
| | | | |
Db 766 TCAAGATATTCAACATTCGCTTTTGGAGAGGTATGTTTTCGTTAGAGGGTTTGACAAAAA 825
| | | | |
QY 6596 gacaagagcacgcg 6608
| | | | |
Db 826 AGAGAGACCCCG 838
| | | | |

RESULT 14
AW737153
LOCUS AW737153 602 bp mRNA linear EST 18-MAY-2001
DEFINITION EST338580 tomato flower buds, anthesis, Cornell University
Lycopersicon esculentum cDNA clone cTOD2A6 5', mRNA sequence.
ACCESSION AW737153
VERSION AW737153.1 GI:7646098
KEYWORDS EST.
SOURCE tomato.
ORGANISM Lycopersicon esculentum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
Lycopersicon.
REFERENCE 1 (bases 1 to 602)
AUTHORS van.der Hoeven,R.S., Bezzeredes,J.L., Matern,A.L., Holt,I.E., Liang
,F., Hansen,T.S., Craven,M.B., Bowman,C.L., Ronning,C.M., Nierman
,W., Fraser,C.M., Martin,G.B., Giovannoni,J.J. and Tanksley,S.D.
Generation of ESTs from tomato flower tissue, anthesis
Unpublished (1999)
Contact: CUGI
Clemson University Genomics Institute
100 Jordan Hall, Clemson, SC 29634, USA
Email: http://www.genome.clemson.edu/orders/index.html
5 prime sequence.
FEATURES
Location/Qualifiers
1..602
/organism="Lycopersicon esculentum"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone="cTOD2A6"
/clone_lib="tomato flower buds, anthesis, Cornell
University"
/tissue_type="flower"
/dev_stage="anthesis"
/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
XhoI; supplier: Tanksley; Flower buds and flowers were
taken from greenhouse plants (4-8 wks old, TA496). They
were immediately frozen in liquid nitrogen and then
size-separated while remaining frozen."

BASE COUNT 158 a 136 c 151 g 157 t
```


Search completed: September 27, 2002, 07:45:12
Job time: 8976 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 27, 2002, 10:26:26 ; Search time 70.37 Seconds
(without alignments)
2360.927 Million cell updates/sec

Title: US-09-840-743-2
Perfect score: 9089
Sequence: 1 MQSIMDSSAVNATEQND.....PRPLMARLHPASKLNKNT 1729

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues
Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_71.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	3949	43.4	1017	2 T48452	hypothetical prote
2	2863	31.5	555	2 T48453	hypothetical prote
3	1559.5	17.2	1207	2 D84781	hypothetical prote
4	1260	13.9	234	2 T48454	hypothetical prote
5	709	7.8	917	2 T05430	hypothetical prote
6	278	3.1	5327	2 T13564	microtubule-associ
7	246.5	2.7	2897	2 B48666	cell proliferation
8	246.5	2.7	3256	2 A48666	cell proliferation
9	240.5	2.6	2938	2 T30249	cell proliferation
10	233.5	2.6	2218	2 B84683	hypothetical prote
11	232	2.6	259	2 D75275	endonuclease iii -
12	219.5	2.4	1871	2 D96796	probable heat choc
13	217.5	2.4	1490	2 T20513	hypothetical prote
14	217	2.4	2022	2 T48818	glucan 1,4-alpha-g
15	211	2.3	1188	2 T05324	hypothetical prote
16	210.5	2.3	1805	2 A34736	nestin - rat
17	208.5	2.3	3924	2 S37431	ankyrin 2, neurona
18	207	2.3	3331	2 T25410	hypothetical prote
19	206	2.3	3187	2 JC5837	364K Golgi complex
20	205.5	2.3	2562	2 T14266	Xin protein - chic
21	205	2.3	1983	2 T00385	KTAA0624 protein -
22	204.5	2.2	2717	2 A34203	DNA-binding protei
23	197.5	2.2	1145	2 T18235	transcription acti
24	197	2.2	1366	2 B86292	F/H2.12 protein -
25	196.5	2.2	3225	2 I52300	giantin - human
26	196	2.2	3488	2 T34418	hypothetical prote
27	192.5	2.1	2954	2 T14156	kinesin-related pr
28	192.5	2.1	3259	1 A56539	giantin - human
29	192	2.1	1888	2 T14273	zinc finger protei

30	191.5	2.1	2094	2 S33124	tpr protein - huma
31	191	2.1	1742	2 T49451	kinesin-like prote
32	191	2.1	2052	2 C97038	phage-related prot
33	190.5	2.1	2447	2 T16870	hypothetical prote
34	190.5	2.1	4957	2 T03455	ALR protein - huma
35	190.5	2.1	5262	2 T03454	ALR protein - huma
36	190	2.1	1284	1 MNVZAI	A-type inclusion p
37	190	2.1	2464	1 QRMSPI	microtubule-associ
38	189.5	2.1	236	2 S75373	probable DNA-(apur
39	189.5	2.1	1634	2 T26517	hypothetical prote
40	189.5	2.1	1804	2 T34518	nestin - golden ha
41	189.5	2.1	3498	2 T22330	hypothetical prote
42	189	2.1	2453	2 S60234	nuclear receptor c
43	188.5	2.1	1435	2 A37793	erythrocyte-bindin
44	188.5	2.1	1926	2 JC4842	DNA-binding nuclea
45	188	2.1	1593	2 T22028	hypothetical prote

ALIGNMENTS

RESULT 1

T48452

hypothetical protein T32M21.160 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 20-Apr-2000

C:Accession: T48452

R:Bevan, M.; Terry, N.; Ardiles, W.; Buysshaert, C.; Dasseville, R.; De Clerck, R.;

ewes, H.W.; Rudd, S.; Lemcke, K.; Mayer, K.F.X.

submitted to the Protein Sequence Database, March 2000

A:Reference number: T24487

A:Accession: T48452

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1017 <BEV>

A:Cross-references: EMBL:AL162875

A:Experimental source: cultivar Columbia; BAC clone T32M21

C:Genetics:

A:Map position: 5

A:Introns: 167/1: 874/1

A>Note: T32M21.160

Query Match 43.4%; Score 3949; DB 2: Length 1017;

Best Local Similarity 99.9%; Pred. No. 1.1e-206;

Matches 755; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MQSIMDSSAVNATEQNDGSRQDVLEFDLNKTPOQKPSKRKRKFKMPKVVVEGKPKRKP 60

Db 235 MQSIMDSSAVNATEQNDGSRQDVLEFDLNKTPOQKPSKRKRKFKMPKVVVEGKPKRKP 294

QY 61 RPAELPKVVVEGKPKRKAATQEKVKSTGSAKKNLKESATKKPANVGDMSNKP 120

Db 295 RPAELPKVVVEGKPKRKAATQEKVKSTGSAKKNLKESATKKPANVGDMSNKP 354

QY 121 EVTLKSCRKALNFDLENPGDARQDSESEIVQNSGANSFSEIRDAIGGTNGSLDSVSQ 180

Db 355 EVTLKSCRKALNFDLENPGDARQDSESEIVQNSGANSFSEIRDAIGGTNGSLDSVSQ 414

QY 181 IDKTNGCLAMNQPLEVSGNMGNDPKLSTGAKLARDQDPDLTRNQOCQFPVATQTFPME 240

Db 415 IDKTNGCLAMNQPLEVSGNMGNDPKLSTGAKLARDQDPDLTRNQOCQFPVATQTFPME 474

QY 241 NQAAWLQMKNLQIGPFPGNQPMWIRNQOCLANGNQPMWIRNQOCLANGNQPMWIRNQOCLANGNQ 300

Db 475 NQAAWLQMKNLQIGPFPGNQPMWIRNQOCLANGNQPMWIRNQOCLANGNQPMWIRNQOCLANGNQ 534

QY 301 PQGNKRPIFLNHTQCLPAGNQLYGSPTDQHLVMSGTGQGHLLIKNQPGSLIRGQQPC 360

Db 535 PQGNKRPIFLNHTQCLPAGNQLYGSPTDQHLVMSGTGQGHLLIKNQPGSLIRGQQPC 594

QY 361 VPLIDQQPATPKGFTPLHNMVATSMSPGLRPHSOSQVPTTYLHVESVRLNGTGTGQC 420

|||||


```
Db 332 EGIFQGRQESLNVLSKIDTPIKKTTGTHARFNLSSMKNLV-----EV 376
QY 399 PTTLHVESVRIKNGTGTQCRAPAYDSLOODIHQGNKYILSHSINGCKKALPQ 458
Db 377 PE---HL-----TSYGCKPQ-----QNNKILVDTRVT-----401
QY 459 NSSLPPTIMAKLEBARSKROYHRAMGQTEKHDLNLAQIAQSOODVERHNSSTCVELDA 518
Db 402 -----VSKKPKTKSEKQ-----414
QY 519 AKTKIOKVOENLHGWPPEVIEIEDPTDGARKKNTASIKASKGNSSPVKKAETKE 578
Db 415 ---TK-OKNLLPNLCRPPSTGLSPD-----437
QY 579 KCIVPKTPAKKGRAGKKSVPVPAHASEIQWLQWPTPLRSKPKGKGRKSIQDSGKA 638
Db 438 -----ELW-----KRNST-----446
QY 639 RCPGSELLCQDSIAEIIYRMONLYLQKEREQONAMVLYK-----GDGALVPE 688
Db 447 -----ETISEL-----LRLLDINREHSETALVPVTMNSQIVLFGGAGAIVPVT 490
QY 689 S-KKKRPPKVIDDETRINWLLMGDEKEGDEEDKKKEKWEEERRVFRGRADSEFI 747
Db 491 PVKRPKRPKVLDODETRVWKKLL-ENINSEGVGGDEQKAKWEEERNVFRGRADSEFI 549
QY 748 ARMHLVQDRRFSPKGSVSDVIGVFLTONVSDHLSSAFMSLAARFP-PKLSSSREDE 806
Db 550 ARMHLVQDRFETPKGSVSDVVGVLTONVSDHLSSAFMSLASQFPVPFSSNFD- 608
QY 807 RNRSVVVEDPEGCLINLNETPSWQ-----EKVQHPSDMEVSGVDSGSKQLRDCSNS 859
Db 609 -----AGTSSMPSTQITLYDSEETMSPPDHNSVT-----640
QY 860 GIERENFLEKSIQNLBEEVLSSQDSFDPAIFQSGRVCSCSKSDAEFTTRCET--KT 917
Db 641 -----LKNTPQDEEKDYVPSNET-----SRSSSEIAISAHESVDKT 676
QY 918 VSGTS--QSVQTGSPNLSDEICLOGNERPHLYEGSGDVQK-QETTNVAQKKPDLEKTMW 974
Db 677 TDSKEYVSDSRKGS-----SVEVDKTEKCRVNLNLPSEDSALTC 716
QY 975 KDSVCFQCPNRDNTWQTPSSSYEQCATROPHVLIDIEFGMQGELGYSWMSISPRVDRV 1034
Db 717 QHSMVSDAPQNT---ERAGSS-----EIDLEGF-YRTSEMKL-----750
QY 1035 KKNVPRFRFGGSGVPRFTGTIIPSTPHELPGMGLSGSSSAVQEHODDTQHQQDEMN 1094
Db 751 -----LQGVQVSLDSNQVSPN-----MSPGDCSSEIKGFQ-----SMKEPT 787
QY 1095 KASHLQKTFDLNLSSEE-CUTROSSTKQNTDGCPLPRDRTAEVDVDPPLSNSSLQNLV 1153
Db 788 KSS-----VDSSEPGCCSQ-----DG-----DVL-----807
QY 1154 ESNSNKEQTAVEYKFNATILREMGTLADGKKPTQOWSLRKDXDVEGNEGQRNKNM 1213
Db 808 ---SCOKPT---LKERGKKYLKE-----EKKAEDWDCLREAOARAGIREKTRSTM 852
QY 1214 DSIDYEAIRRASISFISPAIKERGMNMLAVRI-----KDFLERIVKDHGGIDLE 1263
Db 853 DTVDMKAIKRAADVKEVATIKHSMNKLAEIRIQLYTLNMKIMQGFLDRLVNDHGSIDLE 912
QY 1264 WLRESPPDKADYLLSIRGLKSVCEVRLTLHLNLAFFPVDNTNGRIAVRMGWVPLQPL 1323
Db 913 WLROVPPDKA-----923
QY 1324 ESQLHLELVPVLESQKFLWPLRCKLDORTLYELHYQLTTFKGVCTKSRPNCNACPM 1383
Db 924 -----YELHYQMITEFGKVFCTKSKPKNCNACPM 950
QY 1384 RGEGRHFASAVASARLALPAPEERSLTSATIPVPPPEFPVIAIPMIELPPLPKSLASGA 1443
Db 951 KGEGRHFASAFASARLALPSTEKGMGTPDKNPLP-----LHLPEPFQREQGSEV 999
```

```
QY 1444 PSNREN-----CEPIIEEPASPCQECTEITESDIEDAYYNEDPDEIPTIKLNIEQFGMT 1497
Db 1000 VQHSEPAKKVTCCEPIIEEPASPEPEAEVSIADIEEAF-EDPEEPTIRLNMDAFTSN 1058
QY 1498 LREHMERMELOEGDMSKALVALHPTTTSIPTPKLKNISRLRTEHQVYELPDSHRLDGM 1557
Db 1059 LKIMEHNKLDGDNMSSALVALTAETASLPMPLKNISQLRTEHRYVELPDEHPLLAQL 1118
QY 1558 DKREPDDPSYLLAIWTPGETANSAPPEQKCGKASGKMGCFDETCSECNSLREANSQTV 1617
Db 1119 EKREPDDPCSYLLAIWTPGETADSIQPSVSTCIFQANGMLCDEETCFSCNSIKETRSQIV 1178
QY 1618 RGTILLI 1623
Db 1179 RGTILV 1184
RESULT 4
T48454
hypothetical protein T32M21.180 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 20-Apr-2000
C:Accession: T48454
R:Bevan, M.; Terry, N.; Ardiles, W.; Buysshaert, C.; Dasseville, R.; De Clerck, R.;
ewes, H.W.; Rudd, S.; Lemcke, K.; Mayer, K.F.X.
submitted to the protein Sequence Database, March 2000
A:Reference number: Z24487
A:Accession: T48454
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-234 <BEV>
A:Cross-references: EMBL:AL162875
A:Experimental source: cultivar Columbia; BAC clone T32M21
C:Genetics.
A:Map position: 5
A:Introns: 49/2; 61/3; 81/1; 127/3; 149/3; 188/1; 201/1
A:Note: T32M21.180
Query Match 13.9%; Score 1260; DB 2; Length 234;
Best Local Similarity 100.0%; Pred. No. 1.9e-61;
Matches 234; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1496 MTLREHMERNELOEGDMSKALVALHPTTTSIPTPKLKNISRLRTEHQVYELPDSHRLD 1555
Db 1 MTLREHMERNELOEGDMSKALVALHPTTTSIPTPKLKNISRLRTEHQVYELPDSHRLD 60
QY 1556 GMDKREPDDPSYLLAIWTPGETANSAPPEQKCGKASGKMGCFDETCSECNSLREANSQ 1615
Db 61 GMDKREPDDPSYLLAIWTPGETANSAPPEQKCGKASGKMGCFDETCSECNSLREANSQ 120
QY 1616 TVRGTLILPCTAMRGSPFLNGTYFQVNEFLFADHESLSKPIDVPRDWLDPRTVYFGT 1675
Db 121 TVRGTLILPCTAMRGSPFLNGTYFQVNEFLFADHESLSKPIDVPRDWLDPRTVYFGT 180
QY 1676 SVTSIFRGLSTEQIOQCFWKGFCVCRGFEQKTRAPRPLMARLHFPASKLNKNT 1729
Db 181 SVTSIFRGLSTEQIOQCFWKGFCVCRGFEQKTRAPRPLMARLHFPASKLNKNT 234
RESULT 5
T05430
hypothetical protein F28A23.180 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 24-Nov-1999
C:Accession: T05430
R:Bevan, M.; Weichselgartner, M.; Fartmann, B.; Granderath, K.; Dauner, D.; Herzl, A.
submitted to the protein Sequence Database, October 1998
A:Reference number: Z15415
A:Accession: T05430
A:Molecule type: DNA
A:Residues: 1-917 <BEV>
```

A:Cross-references: EMBL:AL021961
A:Experimental source: cultivar Columbia; BAC clone F28A23
C:Genetics:
A:Map position: 4
A:Introns: 41/3; 70/3; 88/1; 118/1; 321/1; 367/1; 474/1; 504/2; 624/3; 655/2; 667/3; 695/3
A:Note: F28A23.180
C:Superfamily: Arabidopsis thaliana hypothetical protein F28A23.180

```
Query Match          7.8%; Score 709; DB 2; Length 917;
Best Local Similarity 21.4%; Pred. No. 1.2e-30;
Matches 232; Conservative 108; Mismatches 163; Indels 580; Gaps 22;

QY 651 IAEIIRYQNLVGLGDKEREQONAMVLYKGDGALVPYESKKRKPDKIDDETRIWNL 710
DB 388 IAKLIKMGRLKINK-----VTTMIKADKLV-----TAKVNLDPETIKENDV 431

QY 711 LMGKDEGEDEEDKKKEKWEERRVFRGRADSFIAHMLVOGDRFPWKGSVDSV 770
DB 432 LM--VNDSPRSYDDKETEAKWKEREIFQTRIDLFINRHLQGNRRKFKQWKGSVDSV 489

QY 771 IGVFTQNVDSHLSSAFMSLAARPPKLSRSSREDERNVRVSVVDEDPGCGILNLEIPSW 830
DB 490 VGVFTQNTDYLSSNATMSVAAKFP-----VDAREGLSYIEEPQ----- 530

QY 831 QEKVOHPDMEVSGVDSGSKQELRDCNSGIERFNFLKSIQNLEEEVLSSQDSFPAIF 890
DB 531 -----DAKSSSECI----- 538

QY 891 QSCGRVSGSCSKSDAEPTTRCTKTVTSQSVQTSQSPNLSDIEICLOGNERPHLYEGS 950
DB 539 -----ILSDE----- 543

QY 951 GDVQKETTNAQKPDLEKTMNWKDSVCFGQPRNDTNMQTPSSSYBQCACATRPHVLDI 1010
DB 544 ----- 543

QY 1011 EDFGMQGEGLGYSNMISPRVDRVKNKNVPRFRQGGVSPREFTGQIIPSTHELPGMG 1070
DB 544 ----- 543

QY 1071 LSGSSAVOEHOQDQHNQOQDEMKNKASHLQKTFDLNLSSECLTRQGSTKQNTDGLCP 1130
DB 544 ---SISKVEDHEN-----TAKRKNKKTGII- 565

QY 1131 RDRTAEDVVDPLSNNSLQNLIVESNNSNKEQTAVEYKETNATILREMKGLADGKKPTS 1190
DB 566 -----EDEIVD----- 571

QY 1191 QWLSLRKDVEGNEGRQERNKNMDSIDYEAIRRASISEISAIKERGMNMLAVRIKDFL 1250
DB 572 -WNNLRR-MYTKEG--SRPEMHMDSVNNVSLGQNVLETTIKKRGQFRILS----- 620

QY 1251 ERIVKHGGIDLEWLRSPDPKAKDYLLSIRGLGLKSVCEVRLTLHLNAPVDTNVGRI 1310
DB 621 ERIL-----VDITNVGRI 632

QY 1311 AVRGMVPLQPLPESLQHLLELYPVLESIQKELWPLRCKLDQRTLYELHYQLITFGKVF 1370
DB 633 AVRGLVPLEPLPQNGVQH-----QLFEYELHYOMITFGKVF 669

QY 1371 CTKSRPNACPMRGCEHRHFAAYASARLALPAPPEERSLTSATIPVPPEP-----PPVA 1425
DB 670 CTKTIPNACPMKSECKYFASAVSSKVLLESPEEK-----HEPNTFMNAHSQDVA 722

QY 1426 IPMTLEPLPLEKSLASGAPSNCEPIIEEPASPGOECEITEISDIEDAY-----NEDPD 1482
DB 723 VDMTSNINLEECVSGSGDQAICYKLPVEFPSSPRAEIPEST--DIEDVPFMNLYGSA 780

QY 1483 EPTIKLNIQFGMTLRHM--ERNMELQEGDMSKALVALHPHTTSTPTP---KLKNISR 1537
DB 781 SVPKIDFDLALKSVEDALVISGRMSSSDEEISKALVITPTPENACIPIKPPRKMKNYNR 840
```

```
QY 1538 LRTEHQVYELPDSHRLDGMCKREPDDPSPYLLAIWTPGETANSAQPPEQKCGKASGKM 1597
DB 841 LRTEHVYVLPDNLHELHDFERRKLDLDDPSVLLAIWQP----- 878

QY 1598 CFDETCSECNLSREANSQTVRGTLILPCRTAMRGSFPLNGTYFOVNFELFADHESSLKPID 1657
DB 879 ----- 878

QY 1658 VPRDWINDLPRTVYFCTSVTSIFRGLSTBOIQCFWKGFCVCGVEQKTRAPRLMARL 1717
DB 879 -----GFICLRAFDRKQKORDPKELVRL 900

QY 1718 HFP 1720
DB 901 HTP 903

RESULT 6
microtubule-associated protein homolog - fruit fly (Drosophila melanogaster)
N:Alternate names: hypothetical protein EG:49E4.1
C:Species: Drosophila melanogaster
C>Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 17-Nov-2000
C:Accession: T13564
R:Spanos, L.; Papagiannakis, G.; Siden-Kiamos, I.; Louis, C.
submitted to the EMBL Data Library, April 1999
A:Description: Sequencing the distal X chromosome of Drosophila melanogaster.
A:Reference number: Z17689
A:Accession: T13564
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-5327 <SPA>
A:Cross-references: EMBL:AL031128; PIDN:CAA20006.1
C:Genetics:
A:Cross-references: FlyBase:FBgn0025392
A:Introns: 24/2; 52/3; 104/3; 179/1; 232/1; 1669/3; 2566/1; 4798/3; 5272/1
A:Note: EG:49E4.1
C:Superfamily: Drosophila 576K microtubule-associated protein homolog
```

```
Query Match          3.1%; Score 278; DB 2; Length 5327;
Best Local Similarity 19.7%; Pred. No. 4.3e-06;
Matches 277; Conservative 198; Mismatches 583; Indels 348; Gaps 59;

QY 2 QSIMDSAVNATEATEQDNGSRQDVLEFLNKTPOQPSKRKRKFMKPVVVEG-----K 55
DB 1986 ESID-----EAKSKEESRRE-----SVAESPLPSKEASRPASVAESTKDAEK 2031

QY 56 PKRPRKPAELPKYVVEGPKRKP-----RKAATQEKYKSKETGSAKKNLKESATKK 108
DB 2032 SKEESRRESVAESKPLPSKEASRPASVAESIKDEAKSKEESRRESAAKSPKPSKEASR 2091

QY 109 PANVG---DMSNKSPEVTLKSCRKALNFDLENDGDAR--OGDSE--SEIVONSSGANSF 160
DB 2092 PASVAESVKDEADKSKEESRRE-----SMAESGKAQSIKGDQSPKLEYSRPSVAESV 2144

QY 161 SEIRDAITGGTNGSFLDSVSDIKDTNGLGMNQPLEVSMGNQPKLSTGAKLARDQDPDLL 220
DB 2145 KD--DPVKSKEPSRRESVAGSVTADSDQSPLESKAGRPESVVDVSKDEAKQESRR 2202

QY 221 TRNQOCCOFFVATONTOPPMENQQAWLQMLNQLIGFPFGNQOQPRMTIRNQOCPCLAMGNOQP 280
DB 2203 ESKTESVIPPKAKDKSPKEVLQ-----PVSMTET--TIREDA-----DQP 2240

QY 281 MYLIGTRPALVSGNQ-----LGGPQGNKRPIFLNHQTCPLPAGNQLYGSTPDMHOLV 333
DB 2241 M-----KPSQAESRRESIAESIKASSPRDEKSPSLASKEASRPGSVAESIKYDLDKPQII 2294

QY 334 MSTGGQOH---GLLIKN-----QPGSLIRGQQPCVPLID--QQPATPK-----GF 374
DB 2295 KDKSTEHRSRESLEDKSAVTSKESVSRPLSVASDHEAAVAIEDDAKSSISPKDKSRPGF 2354

QY 375 THLNQVATSNMSSPGLRPHSQSQSVPTTYLHVESYSR-----ILNGTTGTQCRSRAPAYD 428
```


Qy	364	IDQOPATPKGFTHLNOVATSMSSPG-----LRPHSQSQVPTTYLH--VESVSRILNGTTG	417
Db	1562	LFOTPASGDPVTVDETTHIALQSPQPGHIIINPASMKRQSNMSLRKOMREFSILEKOTQS	1621
Qy	418	TCORSRAPAYDSLOQDHOGNKYILSHE-----ISNGCKCK--ALPONSSSLPTPIMAK	469
Db	1622	RGDRAGTPA--PMQEB--NGTTAIMETPKQLDFTGNSTGHKRPRTPKNAQP-----	1671
Qy	470	LEEARGSKRQYHRAMGOTEBKHLNLAQOI--AQSQ-DYERHNSSTICEVYLDAAKKTQIK	526
Db	1672	LEDLDGFQELFOFPAGASDPVSVEESAKISLASSQAEPVRTPAST-----KRRSKTGLSK	1726
Qy	527	VVOENLHMPPEVIEITDDPTGARGKNTASISKGASKNGSSPVKKTAEKECIVPKTP	586
Db	1727	V-----DYRQEPETGLGRMK-----SILGRAPGTPAPQOE--ENDSTAFMETP	1766
Qy	587	AKK-----GRAGRKKSVPPIAHASEIQLWQPTTPKPTPLSRSKPKGKRKSIQD---SGK	637
Db	1767	KQLDFTGNSGHKRR-----POTPKIRAQP-----LEDLDGFOEL	1802
Qy	638	ARGPSGELLCDQSIABEIIYRMONLYIGDKEREQEQNAMVLYKGCALVPYESKK--RKPR	695
Db	1803	FQTPAG-----ANDSVT-----VEESVKMSLESSQAEPVKTPASTKRLSTGL	1845
Qy	696	PKVIDIDETTRINWLMGDKEGDEEKDKKKKEWDEERVRVGRADSTIAR-----	749
Db	1846	SKYVDREDPSILEKTKSPGTPAPVQIENDCDTAFMETPKQKLDFTGNSGHKRRPRTPKI	1905
Qy	750	---MHLVGODRRFSPWKSSVVDVIGVFLTONVSDHLSSAFASWLAARPPKLESSRED	805
Db	1906	RAQPLEDLGFOELFOFPAGASDSV-----TVEESAKMSLESSQAEPVKTPASTKRLS	1958
Qy	806	ERNVRSV--VVEDPEGC---ILNLNIEIPSMQEKVQHPSD-----MEVSGVDSSGSK	850
Db	1959	KTGLSKVDREDPSTLGGKTKSPGRAGTPAPVQIENDSTAFMETPKQKLDFAENSSGSK	2018
Qy	851	EQLRDCNSGIERPNFLEXSIONLBEVLSSODSPDPAIFOSGCRVSGCSKSDAEFFT	910
Db	2019	RRSRTSKN---RSOPLE--DLBGFOELFOFPAGASNVPVSVEESAKI---SLESSQAEPVR	2070
Qy	911	TRCETKTVSGT---SQSVOTGSPNLSDEICL-----QGN-----	942
Db	2071	TRASTKRLSKTGLNKMVDREGHSPUSKSSCASQKVNQTLTGEDHGRETKDGKVVLLAQKL	2130
Qy	943	RPHLYEGSGDVOKQETTNVAQKKPL-----EKTMMKDSVCFGOPRNDTNQWTPSS	995
Db	2131	EPALIVTRGKRQORSCKKSQSPEDLSGVQEVFOFSGHNKDSVIV-----DNLAKLPSS	2184
Qy	996	SYEQCATRQPHVLIDEDFQMOGEGGLGYSNMSSISPRVDRVKNKNVPRFFROGGVSPREFT	1055
Db	2185	S-----PPLEPTDTSVTSRRQARTGL-----RKHVKN-----LSGGIMHPQIS	2224
Qy	1056	GOIIPSTPHLPQMG--LSGSSSAVQEHQDDTOHNOQDEMKNKASHLOKTFLLD-----	1106
Db	2225	GEIV-DLPREPEGEGVKITRKQSVKRLDTEVNVPRSKRQIRIABKTLBUDLPFGFOELC	2283
Qy	1107	-----LNSSECLTRQOSTKONITDGLCPDRPTAEV---	1138
Db	2284	QAPSLYMDSVIVIEKTPKMPDKSPEDVTITSEQARRRLRLVTEEP1PORKTRVVVRQT	2343
Qy	1139	-----VDPLSNNSLQNIL-----VESNSSNKEOTA-	1164
Db	2344	RNTQKEPISDNQGMEEFKESSVOKQDPSVSLTGRNRQPRTVKEKTQPLEELTSFOEETAK	2403
Qy	1165	-----VEYKETNATILREMKGTILA--DGKK--PTSQWDSLRKDVCEGNEGRQRNKNM	1213
Db	2404	RISKSQPOPEKETETLAGLQRIOLINDGVKEEPTAQ---RRQ-PSRETRNTLKPEVG	2458
Qy	1214	DSIDYEAIRRASITSEISEAIKERGNNMLAVRIKDFLERIVKDHG-GIDLEWL-----	1266
Db	2459	DSINVEEVKSKTKQKI-----DPVASVPVKRRPRVPKEKAQALELUGLGPOTL	2509
Qy	1267	-----ESPDK-----AKDYLLSTRGLGKSV-	1289

Db 2510 GHTDESASDKGPTQMPNCNSLQPEQVDSFOSSPRPRTRRGKVEADEEPSAVR-----KTVS 2565
QY 1290 CVRLLLTHLNAFFVDNVG-----RIARVMGW--VPLQLPLPE 1324
Db 2566 TSR-OTWRSRKVPDEIGNNGTQVSKASIKOTLDTVAKVTSRRQLTHKGWGSTLLKLLGD 2624
QY 1325 SLQL-----HLLELYPLVESIQK---FLWP-RLCKLDQRTLYELHYLIITFGKVFC 1371
Db 2625 SKETQISDHSEKLAHDTLSLKSTQOKPDSVKPLRTRCRVLRSASKEVPKVELVDTRDHA 2684
QY 1372 T-KSRPNCNACPMRGCRHFASAYASARLAL-----PAPERSLTSATTVPVPPES 1420
Db 2685 TLOSKSNPLSPRRKSGARDGSIVTRRALSLAPKQEADEKVPVPEKKAASSKRYVSP- 2743
QY 1421 FPPVAIPMIPLPLEKSLAS-----GAPSNRECEPIIEPASPQOECTEITSDIEDA 1475
Db 2744 --PVKKMKHLKI-----VSNKLESVEEQVSTYMKTEWEAKRENVPVTDQ-----NS 2787
QY 1476 YYNEDPDEIPTKILNTEQFGMTLREHMERNMELQEGDMSKALVALHPTTTSIPTPKL--- 1532
Db 2788 RYRK-----KTNVQK-----PRKPKFAS 2805
QY 1533 -----KNISRLRTEHQHYVELPDSHRLLDGMDKREPDDPSPYLLAIWTPGETANSAQPP 1585
Db 2806 AENVGIKKNEKTKMTASQETEL-----QNPDD-----GAKKSTSR- 2840
QY 1586 EQQKGKASKCMGFDETCSECNLSREANSQTVRGTLILI--PCRTAMRGSGFP 1634
Db 2841 ----GOVSK-----RTC-----LRS-----RGTEMPQPCAEKTSKP 2871

RESULT 10
B84683
hypothetical protein At2g28300 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C:Accession: B84683
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.;
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana
A:Reference number: B84420; MUID:20083487
A:Accession: B84683
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-2218 <STO>
A:Cross-references: GB:AE002093; NID:g4803953; PIDN:AAD29825.1; GSPDB:GN001340
C:Genetics:
A:Map position: 2

—

Qy	1200	EGNEGROERNKNNMDSIDYEATRRASISEISAIEKRGMMNMLAVRIKDFLE--RIYKDH	125
Db	1156	MEQGFSETHARTDS---GGIDRG--NEVSENMSDGVKYMITSISVVQVPDASHDLNVSDQ	1210
Qy	1258	-----GGIDLEWLR-----SPPDKAKDYLLSIRGLGLKSVCEVR---LLTLHLAFPV	1303
Db	1211	TDPLVGGIDPEHVQENVDPASPASPHGAAANVIVFQSEGLSPSILPDDVACQLESMSNDE	1270
Qy	1304	DTNVGR-----IAVRMGWVLPQ-----LPESLQLHLLELYPV-----	1336
Db	1271	KTNISSQVDPVSHDLKVSDQTDIPVGGIVPENLQ-----EIVDPASPAGVVPDVVVS	1326
Qy	1337	-LSIQKFLWPRCLKLDQRTLYELHYQLITFGKVFCTKSRNCNACPMRGECRHFA--	1392
Db	1327	QSEEIQS---PSILPD-----VPGQPDGNGCKMDMTQMNTSIDI	1364
Qy	1393	AYASARLALPA---PEERSLTSATIPVPE-----SFPVPAIPMIEPLPLEKSLASG	1442
Db	1365	GITSKTCQPSSTQPEDENRNSLSHCEPSEVVEQRDSRDQVCSVESQVEISALEN	1424
Qy	1443	APSNRENCEPI-----IEEPASPGQCEFTI-----ESDIEDAYYNED	1480
Db	1425	RSADIOQPQSILVDQDIEESKEPGIESADVSLHQLADIQAEPNSLVDMQMDIES--KEP	1482
Qy	1481	PDEIPTKL-----NIEQFGMTLREHMERNMELQEGMSKALVALHPTTTSIPTPKLNIS	1536
Db	1483	GTESADVSLHQLADIQGPSILVDQMDTEKSEPGTES-ADVSLHQLADIQGPSIL-VD	1540
Qy	1537	RLRTE-----HQVVEL-PDSHRLLDGMDKREPDDP--SPYLLAIWTPGETA	1579
Db	1541	QMDTEKSEPGTESADVSLHQLADIQGPSILVDQMDTEKFPNDVSLHQLADIEPSLSI	1600
Qy	1580	NSAQ 1583	
Db	1601	SAVQ 1604	
RESULT 11			
D75275			
endonuclease III - Deinococcus radiodurans (strain R1)			
C:Species: Deinococcus radiodurans			
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Mar-2000			
C:Accession: D75275			
R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; D.			
M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zal.			
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.			
Science 286, 1571-1577, 1999			
A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiod			
A:Reference number: A75250; MUID:20036896			
A:Accession: D75275			
A>Status: preliminary			
A:Molecule type: DNA			
A:Residues: 1-259 <WHI>			
A:Cross-references: GB:AE002073; GB:AE000513; NID:96460244; PIDN:AAF11977.1			
A:Experimental source: strain R1			
C:Genetics:			
A:Gene: DR2438			
A:Map position: 1			
Query Match 2.6%; Score 232; DB 2; Length 259;			
Best Local Similarity 31.4%; Pred. No. 1.7e-05;			
Matches 60; Conservative 30; Mismatches 81; Indels 20; Gaps			
Qy	1217	DYEAIRRASISEISAIEKRGMMNMLAVRIKDFLERIVKDHGGIDLEWLRESPPDKAKDY	1276
Db	79	DWAIDAITLAPDAVAHAIRRSNYPESKAPRIQETLRRKIAAFPGYDLDPLRDPVKDAKW	138
Qy	1277	LLSTRGLGLKSVCEVRLTLHLNLAFVDPDINVGRIARVWGPVLPQLPESLQLHLELYPV	1336
Db	139	LTDLPVGGVKTASLVLLFNVARVPFVDPDTHVRHSVTRGVCIPIRMG-EQAHALLALLPP	197
Qy	1337	LESIQKFLWPRCLKLDQRTLYELHYQLITFGKVFCTKSRNCNACPMRGECRHFA--	1396

Db 198 -----DPPYLYELHINFLSHGRQVCTWTRPKGKCILLRRC-----DAYAL 238
QY 1397 ARLALPAPEER 1407
Db 239 YGDKVPFSEK 249
RESULT 12
D96796
probable heat shock protein, 53413-59028 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
R;Accession: D96796
C;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Liu, Y.; Liu, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A;Reference number: A86141; MUID:21016719
A;Accession: D96796
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1871 <STO>
A;Cross-references: GB:AE005173; NID:g6143906; PIDN:AAF04452.1; GSPDB:GN00141
A;Gene: F28016.15
A;Map position: 1
Query Match 2.48; Score 219.5; DB 2; Length 1871;
Best Local Similarity 17.28; Pred. No. 0.0014;
Matches 251; Conservative 222; Mismatches 510; Indels 477; Caps 61;
QY 86 EKVKSKETGSAKKNLKESATKPKANVGDMNSKSPVTLKSCRKALNF-DLENPGDARQ 144
Db 103 ERIEEKGLADSNKESVDSSLRKP-----PDTEGRECHQTRHEQENKQLVQA 152
QY 145 DSEIIVNSGANSFSEIRDAIGTNGSFLDSVQIDKTNGLGNAMQPLEVSMGN---- 200
Db 153 ESDD-----SDDFGRFAFEIEE-----QESDVLDRSTSGAMEKEMTDVDDGLRK 199
QY 201 -----QPKLSTGAKLARDQPDLLTRNQOQFPVATQNTQFMENQOAWLQMKNLIGF 255
Db 200 VQTEEPERHNEESKISE-----HVDGETSGHEKKVVKM 234
QY 256 PFGNQOQPRMTIRNQOQCLAMGNQOQPMYLIQTPRAL-----VSGNQOQGGPOGNKRPIFLN 311
Db 235 DKNRDVKEVDG-----AMGEF-----RPNIDRTQVVGDDTAETKNDDEF--- 278
QY 312 HOTCLPAGNQLYGPPTMHQLVMSTGGQOHLKLNQOQPSLIRGQOQPCVPLIDQOQATP 371
Db 279 -----ESDKLEADEVDK-----INEGNK--VRRHSDRLNKLQEK-----EQHSKEQ 322
QY 372 KGFT---HLNQWATSMSSPGLRPHSQSOVPTTVLHVESVSRILNGTTGTQCRSRAPAYD 428
Db 323 KHSKEENKELV-----EKTTP-----EAETIRNDILGPQOEIVPEVD 363
QY 429 SLOODIHQ--NKYILSHSISNGCKALPQNSLSLTPPIA----- 468
Db 364 TLGKTSDEGEKQNVKKEIKNGDATEIDAKMGVEFASNIADTGMNSDEFESDKLESAD 423
QY 469 ---KLEARGSKRQVHRAMQTEKHDLMLAQOIA--QSQDVERHNSSTCVYL-----D 517
Db 424 EVDKMKRQKQNDKVGQASEDLSLKLQEIQEQQFGQKRHDQENIKELREGQASE 483
QY 518 RAK-----KTKIQKVVQENLHGMP-----PEVIE-----IEDPT----- 547

RESULT 13

T20513
hypothetical protein F02E9.4 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T20513

Db 484 AEKNINDILKPVQKRSEGHKIKQTFQEBETNKQPECYNEKIMETGKKINEDGTRKVOEM 543
QY 548 -----DGARKKNTASISKASKNSNPVKKTATKEKCIIVPKTPAKKGRAGRKSVPP 600
Db 544 IRQQELDEPARSEKENRSLVSKTNDEKK---EKE---IAGTERKESDRPKILRE 597
QY 601 PAHASEIQLWQPTPPKTPLSRSKPKGKRKSIODSKARGSPSGELLQCDSIAETIIRMQ- 659
Db 598 QEVADEV-----AEDTKFSIYGEVKEEIEIAGKEKEFGSDDDIARLVRDTEQ 645
QY 660 ---NLVLGDKEREQEQNAMVLYK---GDGALVPYESK-----KRKPRPKVDIDDE 703
Db 646 LDSNAMGQEKQMIQELVLEEKVCDGKGKIIAIAETKAENKSKRVQETEEQKLDKREDT 705
QY 704 TTRIWNLLMGKDEKGEDEKDKKKKWKWEEERVR-PRGRADSFIAHMLVQGDRRRSPW 762
Db 706 CGKHQFKLIEGSDHGEVEDVEKDKRTAEKRIKDRAREAEIEKDKLGVSGRYI--- 762
QY 763 KGSVVDSDVI---GVELTONYSDHLLSSSAFNSLAARFPKLLSSSRREDERNRVSVVVEDEP 819
Db 763 KGTIKELVENRGIYRNEH-----EKKKDDAN-----RPEK 794
QY 820 CILNLEIPSWQEKVQHPSDMEVSGVDGSK-----EQLRDC-SNSGIE-----RPNLE 868
Db 795 ITGTIK-----QELVSLNSQLRQENVEDGKTQELVEKIKDCDEEERGSEESKIKTDVV 849
QY 869 KSTQNLSEFVL-----SSQDSFDPALFQSGRVSQSCSK 903
Db 850 RKVQGIKEEELYPKREHGTKITELVEETTDYKQEKETAESDIEAECG-----SLRK 904
QY 904 SD-----AEPFTR-----CETKTVSGTSQSVOGTSPN 931
Db 905 VDGIEHELHEPKIHKERNRNVGTGAREPSQGEKEKEKIVESMTITENDNSIDVOETK 964
QY 932 LSDIEICLQGNRPH---LYEGSDGVQKQ-----ETTNAQKK----- 965
Db 965 KERPGRLSHDKRYKIQELLMEAGHNDKKEEQENNVTAEVELETERVSSKKVQEGRMED 1024
QY 966 -----PDLEKTMW-----KDSVCFQOPRND-----TNWQ 990
Db 1025 DNGKPFHEFERKSYEDWTHEKREKRVLVEEETYPKDKHTGGEDHNDHKEEQKENVI 1084
QY 991 TTPSSSYEQCATROPHVLDIEDFCMOGEGGLGYSWMSISPRVDRV---KNKNVPRRFRFQ 1046
Db 1085 AKALNTEESFKVVEIEKHODHGLKARSWQAKROETEEDKDKTRAMEKNETVERKQTK 1144
QY 1047 GGSVPREFTQIIPSTPHELPGMGLSGSSSAVQE-----HODDTQHNOQDDMNKASH 1098
Db 1145 DGSF-----GKLRGEGDEPLGGHERRGEEDRIEELVETEISDHKEKVKKDEYILRSQD 1199
QY 1099 LQKTFDLNLSSECLTRQ--SSTKQNTDGLPRDRTAEDVDPLSNSSSLQNLIVESN 1156
Db 1200 TKG--VDLGERERSKQKTHKSVEDEIGD---QEDEDAAEAAAVVSRN-----ENG 1246
QY 1157 SSNKEQTAVE---YKETN-----ATILREMGKTLADGKKKPTSQWDSLR 1196
Db 1247 SSRVQTEIESEKHKQNKIPETSNPEVNEDEERVVEKETKEVEAHVQELEKTENC- 1305
QY 1197 KDVEGNEGROERNKNNMDSIDYEAIRASISTEISEAIKERGMNNMLAVRIKDFLERIVKD 1256
Db 1306 KDDGEGRRERERGGQMTA-----ENMLRQREK-----TKS 1336
QY 1257 HGGI--DLEWLRESPPQAK 1274
Db 1337 DDGIVRKIQETKEEPEDEK 1356

R.Gray, I.
submitted to the EMBL Data Library, November 1996
A:Reference number: Z19285
A:Accession: T20513
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1490 <WIL>
A:Cross-references: EMBL:Z81494; PIDN:CAB04052.1; GSPDB:GN00019; CESP:F02E9.4
A:Experimental source: clone F02E9
C:Genetics:
A:Gene: CESP:F02E9.4
A:Map position: 1
A:Introns: 225/3; 321/2; 373/3; 474/3; 619/3; 664/3; 697/3; 828/3; 868/3; 902/2; 965/2;

Query Match 2.4%; Score 217.5; DB 2; Length 1490;
Best Local Similarity 19.2%; Pred. No. 0.0013;
Matches 315; Conservative 221; Mismatches 555; Indels 547; Gaps 86;

```
QY 281 MYLIGTPRALVSGNQQLGGPGGNRP-----IFL-----NHOTCLPAGNQLYG 324
  || || || || || || || || || || || || || || || || || || || || ||
Db 1 MY---NPPPGGGGNN---GGDSQQQPTNNATLFLQMIQQSQHQHQCN---QQQQQLEL 53

QY 325 SPTDMHOLVMSTGGQGHLLIKNQPGSLIRGOQPCVPLIDQOPATPKGFTHLNQWATS 384
  || || || || || || || || || || || || || || || || || || || || ||
Db 54 QIRDERILIEQRMQH-----QQQONQLQG-----LNQFPNPLGLFQVQAQAA 101

QY 385 MSSPGLRPHSQ-SQVPTTYLHVESYSRILNLTGTCQSRAPAYDSLOODIHQNKYILS 443
  || || || || || || || || || || || || || || || || || || || || ||
Db 102 QAO--FAQAQGSPIP---FHIGS-----PLQFSHPAASALQQ-----QYL-- 138

QY 444 HEISNGCKKALPONSLPTPIMAKLEARGSKR-----QYHRAMGOTEKHDLNLA 495
  || || || || || || || || || || || || || || || || || || || || ||
Db 139 -----LPSHPAITPFARNSEARNIEFIAQEEAANVPRANSQQQS---PLI 183

QY 496 QOIAQODVERHN---SSTCEYLDAAKTKIOKVQENLHGMPEVIEIDDPDGDARK 552
  || || || || || || || || || || || || || || || || || || || || ||
Db 184 RPIPOQALNTQNLSTQAAQILAAHQVQVQVQHQHPTTPPLALPI-----AQO 236

QY 553 GKNTASISKGASKNGSPVKTAKEKICIVKPTPA-----KKGRAGRKSV 598
  || || || || || || || || || || || || || || || || || || || || ||
Db 237 G-----PISNEVPVPPVPATSGCPQREPQOQGGRRQNPGRKKP 280

QY 599 PPPAHASELIQWOPPTKPTPLSRKPKGKRSKIDSKARGPSGELCQDSIAEIIYRM 658
  || || || || || || || || || || || || || || || || || || || || ||
Db 281 EGPRVDEALAYLRVIKSTFSSDVPVYHREIMKDFRAQRIETPDVI--EQVALLYDS 338

QY 659 QNLYLG-----DKE-----REOQNAMVLYKGDGALVPYES 689
  || || || || || || || || || || || || || || || || || || || || ||
Db 339 PELVLGFTFLPTGYRIYLTDPDRKIVFSSPQMPVLLSPDERRARAIEAQAQVGAIEL 398

QY 690 KKRPRPKVDIDDETRINLLMGKDEK-----EGDEE---KDKKKEKWEEERRV 738
  || || || || || || || || || || || || || || || || || || || || ||
Db 399 GSQEGISK---DEDRTIEDEMDKSEKDDVDGDDDEDESGIEDKNNEEMEDNHL 454

QY 739 FRGRADSFIAHVLVQDRRSPKSGVSDSVIGVFLQTVSDH-----LSSSAFMSLA 792
  || || || || || || || || || || || || || || || || || || || || ||
Db 455 IE-----EITCDRK---KDDCEQOQETMSSELAHTLNIELKKSF--- 497

QY 793 ARFPKPL-----SSSREDERNRVSVVVEDPEGI-----LNLNEIP---SWQEKV 834
  || || || || || || || || || || || || || || || || || || || || ||
Db 498 AR-PTKLVDFMTFIDFFMSDOQYKDKMEKLRKDDDEDEIEENEKTEVDVDPGSPNAPQEI 556

QY 835 QHPDSMEVSGVDSGSKQLRD-CSNSGIERFNFLKSIQNL---EEEVLSSQDSPDPAIF 890
  || || || || || || || || || || || || || || || || || || || || ||
Db 557 KKPDDIEKK--DSSKNLQIESCS-----DYLVSMLANCCIGEPDLLAATIDPLPYL- 606

QY 891 QSCGRVSCSCSKSDAEPTTRCETKTVSGTSQSVQTSQPNLSDBICIQGNRPHLYEGS 950
  || || || || || || || || || || || || || || || || || || || || ||
Db 607 -----GKLLVNGSDA-----TALKIKTILHFSAT-----NDR----- 633

QY 951 GDVQOQETTNVAQKKPDLEKTMNWKDSVCFQCPQRNDTNWOTTPSSSYEQC--ATROPVHL 1008
  || || || || || || || || || || || || || || || || || || || || ||
```

```
Db 634 NDIPPVNRVNSDVDMDLVKOM-----EKCKMGTKKNEKL 668
QY 1009 DIEDFGMQGEGILGYSWMSI--SPRV--DRVKNKNVPRRFR-----QGSVPREFTG 1056
  || || || || || || || || || || || || || || || || || || || || ||
Db 669 KLVAGQGDGATVEMILKKSRYRILYERLKSRTTPNQLSHLWLVINAYANLDITRE--- 725
QY 1057 QIIPSTPHLPG-MGLSGSSSAVQEHQDDTOHQNODEMKNASHLOKTFLLDLNSEEECT 1115
  || || || || || || || || || || || || || || || || || || || || ||
Db 726 QLI---SELPKIMTSGSD-----LEWILQLLGAKEPKN 758
QY 1116 ROSSTKQNTDGLPRDRTAEDVVDPLSNSSLOILVESNSSNKEQTAV-----EY--- 1167
Db 759 R-----PENDMAVRMKDLPAIQPKRGLDQKMLQV-----KNVEAATVCTGLGSPYRF 808
QY 1168 KETNAT-----ILREKMTGLADGKKTOWSDSLRKDVEGNEGRQERNKNMDSIDYEA 1220
  || || || || || || || || || || || || || || || || || || || || ||
Db 809 KDKATDCSGRVELDDDLKGVLD-----TWTSPSSSEDTGSOAKKSNLEFEHKT 862
QY 1221 -IRRASISEISAIEKERMNNMLAVRIKDFLERIVKDHGGIDLE-WLRESPPDKAKDYLL 1278
  || || || || || || || || || || || || || || || || || || || || ||
Db 863 EDERVELDIIVDS--NRTVIEQLSKTLRDYEAMSDKSKFKLDKWLNAS----- 911
QY 1279 SIRGLGLKSVCEVRLTLNLNLAFFVDVTVGRITAVRGVWVPLQPLPESLQIH-LLELYPVL 1337
  || || || || || || || || || || || || || || || || || || || || ||
Db 912 -----RSTTIRVLA-KVFTN-----SAQDFIDAAOKNPLVGLRRIL 946
QY 1338 ESIQK--FLWPRCKLDORTLYE-----LHYQLITFG-KVFTKSRPNCN--- 1379
  || || || || || || || || || || || || || || || || || || || || ||
Db 947 ESLKEKDLLWSRFQODTNRWDALDKOMSAATTILNQHNKYDQAF--KSKPLVNQIE 1004
QY 1380 -ACPMRGSCRHFASAYASARLALPAPEERSLTSATIPVPPSPFPVPAIMIELPLPLEKS 1438
  || || || || || || || || || || || || || || || || || || || || ||
Db 1005 QICEER--RKNNSTDTSPHLILEYTPERKVYRVNDVTGHFFHDL----- 1047
QY 1439 LASGAPSNRENCEPII-----EAPASPOGCEITEIESDIEDAYNEDPDEIPTIK--LNI 1491
  || || || || || || || || || || || || || || || || || || || || ||
Db 1048 --SGTKCDORTKIVLFRILMEWJCQEQG---VOIDL-----DNGETKFGQDLE 1094
QY 1492 EQFGMTLREHMERNNELQF--GDMSKALVALHPTTTSIPTPK-----LKNISRLRTEHQ 1543
  || || || || || || || || || || || || || || || || || || || || ||
Db 1095 DENLMTL-----LNMDGRRICGD-----RVVPVSTLESNESSIDHFSENLHOKRTRT 1143
QY 1544 VYELPDSHRLLDGMDKREPDPSPYLL-----AIWTPGETANSAQPP 1585
  || || || || || || || || || || || || || || || || || || || || ||
Db 1144 FY-----GDSVYMIIRYHHMIQERFAKILSTQAIYAQEHFDNQKKNK 1186
QY 1586 --EQKCGKASKKCMCFDE-----TCSECSLREANSQTVRGTL-----LIP 1624
  || || || || || || || || || || || || || || || || || || || || ||
Db 1187 RWEDGIGADMGRKALQENIKORRAAVNDIRNVRSCPSSTYETTLRELKQLGNAQMDIVA 1246
QY 1625 CRTAMRGSPFLNGTYF-QVNELFAD-----HESSLKPIDVPRDWIMDLPRRTVVF 1673
  || || || || || || || || || || || || || || || || || || || || ||
Db 1247 FEEAVKNLFGPDIVLFNNIDKLFSSLAKNIHATCAERENPIKL-----YLYKRYLOEY 1300
QY 1674 GTSVTSIFRGLSTEOIQF 1691
  || || || || || || || || || || || || || || || || || || || || ||
Db 1301 GQTAEVLRGKNTYRFEF 1318
```

RESULT 14

T48818

glucan 1,4-alpha-glucosidase related protein [Imported] - Neurospora crassa
N:Alternate names: protein 68B2.20

C:Species: Neurospora crassa

C>Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 05-May-2000

C:Accession: T48818

R:Schulze, U.; Align, V.; Hoheisel, J.; Brandt, P.; Fartmann, B.; Holland, R.; Nyakatu
submitted to the Protein Sequence Database, April 2000

A:Reference number: Z24541

A:Accession: T48818

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-2022 <SCH>

A:Cross-references: EMBL:AL353821; GSPDB:GN00112; NCSP:68B2.20
A:Experimental source: cosmid contig 68B2; strain 74
C:Genetics:
A:Gene: NCSP:68B2.20
A:Map position: 2
A: Introns: 1192/3

Query Match 2.4%; Score 217; DB 2; Length 2022;
Best Local Similarity 17.2%; Pred. No. 0.0022;
Matches 340; Conservative 253; Mismatches 720; Indels 668; Gaps 85;

QY 35 PQQPSKRRKKMPKVVVEGKPKRKRKPAELPKVVVEGKPKRKA-----ATOEKVK 89
Db 4 PKPKDK-----PPSPREKTKITPTNSKNGAPTISPTKSATGATPKSV 50
QY 90 SKETGSAKKNLAKESATKK--PANVGD--MSNKP-----EVLKSKRKALNFDLENPGD 140
Db 51 PPTNLSLGHKKKPELKKPEPSLLGDFLLGRSPQVAAORSASRRKTMSDAQNVRE 110
QY 141 AROCDSESEIV--QNSSGANSFSEIDAIGTNGFGLDSVQIDKTNGL-----GAMN 191
Db 111 ELROEMRAAAVRKLQOPGG-----VDRVKAQKASQAAV-----KAEGLPVPAEDARS 160
QY 192 OPLEVSM--GNOPD-----KLSTGAKLARDQOPDLLTRNQOCQFPVATONTOPFMENQQ 243
Db 161 EPTFAVNLGDDVDEEDRMRIKRWQPKPKPKIET-----VEPKGT-----SGSD 208
QY 244 AWLOMKNLGIFPGNQPRTNRNQPCLAGNQOQPMYLGITPRPALVSGNQOLGGPQG 303
Db 209 STAKEKNKVLG--RMADRDPOVLIKEARP-----RP-----GLPKRIISDDHWMKKQ-- 255
QY 304 NKRPFLNHTCLPAGNQLGSPGTDHQLVMSTGGQOQHGLLIRNQPGSLIRGOQPCVPL 363
Db 256 -----KGKGLAKSP-----VPEAGQ-----PIPKDELKRTA----- 281
QY 364 IDQOPATPKGTHLNQWVATSMSPGLRPHSQSOVPTTYLHVESVRLNGTGTQORSR 423
Db 282 --QNPVSQKKIRDAWAORVELPPPPP--PVARRPTVKTYRHAKT-----GETVTVEDE 330
QY 424 APAYDSLQODJHOGNKYILSHEISNGCKKALPQNSLSLPIIMAKLEEARGSKROYHRA 483
Db 331 -----DDDHHRG-----ALSEPDM-----PRKRSS 351
QY 484 MGQTEKHDNLIAQIAOSQ-----VERHNSSTVEYLDAAKTKIKQVVQENLHGMP 537
Db 352 SGESRSSSNPTQCPNDGIRVSLQKRGSD-----DGIRISPVKA-----RSLPD 401
QY 538 EVIEIEDDP---TDGARKGKNTASIS-----KGAS-----KGNSSPVKK 573
Db 402 DGKVRPGPPVSADSSRSISITVSPSSSGRTPSDRSGRTPPPRGASPPRRASTPPRRA 461
QY 574 TAEKEKCIKPTPAKKGRAKKSV-----PPPAHASEIQLWQPTTP 615
Db 462 STPLRKASTPKPARSDHSAASDDVIEVIVEPESEVSKRSPSPPKR-----RLRSPPPP 517
QY 616 KTLPSRSK-----PKGKRKSIQD 634
Db 518 KRKLPRRSGRGARRKPKRSPPTTATQTEITDTRRPGADKPMPTPRNNGSSGED 577
QY 635 SGKARGPSGELLQOOSTAIEIY-----RMONLYGDKEREQEQONAMVLYKGDALVPYESKK 691
Db 578 SDR-RPPTAAGI--DDLAEIPFGSAFSELEPLRGLHTQRTA-----RQS 620
QY 692 RKPRKVIDDDETRIWNLL-----MGKDEKEGDEKDKKKKWEERRV 738
Db 621 AKPKPKPORNESLKLVPNLVKKVMTGAMEKQOEAEPPTGPKPARIESW----- 673
QY 739 FRGRADSFARHMLYQGDREPSKMGKSVVDSVIGVFLFTQNTYSDHLSSSAFMSLAARFPK 798
Db 674 LNNYDVPFEGMGNPLPPVVAPEPLRVSTPE-----RNSKEKLVDRDLPAHREAPER 725
QY 799 LSSSREDERNRVSVVVEDEPGCILNLEINPSQEKVQHP-----SDMEVSGV 845

Db 726 NSKEKLVDTDVSTHKKARTPEISKRASAEIPRRKERQTPVTHEELRKTSTRSDSSEVTEK 785
QY 846 DSGSKBQLRDCNSGIER-----NFLEKSIQNLEEEVL-----SSQDSFDPAI 889
Db 786 RRSKPTTTPD--SAAGLKRRKSTPRPTLPRRSSGSKVKKPRFDVLKEAFKGESSAHKATPMV 844
QY 890 FQSGRGVSCSKSDAEFTTRCETKTYSGTSQSQTGSPNLSDICLQGNRPHLYEG 949
Db 845 YPSC--ETDVESPEPTEHDLLESRRSPQQRSPDSYKRRSASP-----DRPSRAD 892
QY 950 SGDVQKQETTNVAQKRP--DLE--KTMNMKDSVCFQOP--RNDTNMOTTPSSSYEQACR 1003
Db 893 YSSSEFSTRCPSSRRRPTSDLHDLSTILSEDSREYQEPYKBDTDSVSTVSQS---TVTQ 949
QY 1004 QPHVLIDIEFGMOGELGYSWMSISPRVDRVKNKNVPRFRFROGGSVPREFT----- 1055
Db 950 APE-----EAFQOP-----TPLTREASQOOTSQISRSKSGSLKRLTTTKHSDLSV 994
QY 1056 -----GOIIPSTP-----HELPGMGLSGSSSAVOEHQDDTOHNOQDEMKNASH 1098
Db 995 VLSLPDGGQLVPPSPRSRSIKASRLHRKPSKANDSRVNDLLEEFADDEHYHREL----- 1049
QY 1099 LQKTFLD-----LINSSECLTRQSSTKQNTDGLCLPRORTAEDVVD----- 1140
Db 1050 --KTLVDGVVPLLN-----EFVHGDNVDDADRTDSMAKAVVNMGVALEKLTWYHK 1099
QY 1141 -----PLSN-----NSSLONILVESN 1156
Db 1100 RAPLHDIRLEWLEAVSPVYNNVLDVWRLGFQDLIVNLAPPSGKIDENDSLNALPRNE 1159
QY 1157 -----SSNKQTAVEY-----KETNATILREMKGTLDGCKP 1188
Db 1160 DGVLSYNGERVDVAYLLKRLIRIKWYKFLRAAVVIKTPETEDLLKLYGNLOEKART 1219
QY 1189 TSOWDSLRL-----KQVEGNEGROERNKNMWDSDYEAIRRAISEISEAIKERGMNML 1242
Db 1220 RHREEFARMTDEDANNITDTRARDLRLPLDNVRIDPSQVAAKDFEMDLHSSGQRL 1279
QY 1243 AVRIK-----DFLERIVKHGGIDLEWLRESP-----PKAKDYLLSIR 1281
Db 1280 ECQVELMHRDRVNVPSDMGDILIRDISNKAKPWILLFPVPQVYISARKGESPRSMIYMR 1339
QY 1282 GL--GLKSVCEVRLTLHLNLAFPVDTNNGRIAVRMGVLPQL-----PESLQL--HLEL 1333
Db 1340 GRHNGDEWYELIKLUSTTEVOITDWLGI-----LGSDMPPLSRPKPPMSLIHVSS-- 1392
QY 1334 YPVLESTQKFLWPKLCKLDQRTLYELHYQLITFGKVFCTKSRPNCNACPMRGECRHFASA 1393
Db 1393 -----PKADDLD-----VPIGEPSVIGSYDDDEFPTPSYHKRQASAP 1429
QY 1394 YASARLALPAPEERSLTSATIPV--PPESPPPVAIPMIELPLPLE-----KSLAS 1441
Db 1430 VTS-----YPAPKNKHIPYESVDIFRSQPIITPAYGIESVIPRPLNIRKGSVAVVQPPVELAP 1485
QY 1442 GAPSNRENCEPIIEE--PASPGQECTETESDIEDAYNED-----PDEIPTIKLNE 1492
Db 1486 TTRYKRTSSPKLKHYPHSDIDSETESARSDSESSSSSDELEDDEDDVPTIP----- 1538
QY 1493 QFGMTLREHMERNMELOQDMSKALVALHPTTTSI-----PTPKLNISR----- 1537
Db 1539 --GYSLKQPLQIIVB--ESVVSSENTIAPSQASQVGAQPQGGQTPQERTVQRFVASYW 1594
QY 1538 -----LSTE-----HOVYELPDSHRLLD-----GMDKREPDPDS 1566
Db 1595 SNRKGWREINTEPSRILVFPFGSMEVHMLQETPGNKQAYPLQTSCTSEVDMANREAGV 1654
QY 1567 PYLLAIWTP-----GETA-----NSAQPPEQKCGKSGKMCFEDEC--SECNLSREANSQ 1615
Db 1655 PLICLVLPVVMIRSTRALDLEVRVSPESRSLSDSGMFRFRRAASQDEAKALYEAVHQ 1714
QY 1616 T 1616

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 27, 2002, 12:55:52 ; Search time 33.36 Seconds
(without alignments)
2006.778 Million cell updates/sec

Title: US-09-840-743-2
Perfect score: 9089
Sequence: 1 MQSMDSSAVNATEQND.....PRPLMARLHFPASKLKNKNT 1729

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	246.5	2.7	3256	1	KI67_HUMAN
2	225.5	2.5	2492	1	ATRX_HUMAN
3	225	2.5	2468	1	MAPB_HUMAN
4	219.5	2.4	5430	1	ACFT_HUMAN
5	211.5	2.3	2842	1	APC_RAT
6	210.5	2.3	1805	1	NEST_RAT
7	208.5	2.3	3924	1	ANK2_HUMAN
8	207	2.3	1723	1	AIM1_HUMAN
9	204.5	2.2	2717	1	ZEPI_HUMAN
10	200.5	2.2	3130	1	DPO2_HUMAN
11	192.5	2.1	1781	1	AKAC_HUMAN
12	192.5	2.1	1812	1	BRC1_MOUSE
13	192	2.1	2442	1	CBP_HUMAN
14	191.5	2.1	2349	1	TPR_HUMAN
15	190	2.1	1284	1	ATL_COMPX
16	190	2.1	2464	1	MAPB_MOUSE
17	189	2.1	2453	1	NCRI_MOUSE
18	188.5	2.1	1435	1	EBAL_PLAFC
19	188.5	2.1	3396	1	PGCV_HUMAN
20	188	2.1	2805	1	MAPA_HUMAN
21	187	2.1	3911	1	AKA9_HUMAN
22	186.5	2.1	1132	1	YKKS_YEAST
23	186.5	2.1	1833	1	ZEPI_MOUSE
24	186	2.0	1531	1	NFT5_HUMAN
25	184.5	2.0	1875	1	MLP1_YEAST
26	183	2.0	2441	1	CBP_MOUSE
27	182.5	2.0	1861	1	MAP2_RAT
28	182	2.0	2688	1	ZEPI_MOUSE
29	181.5	2.0	3381	1	PGCV_BOVIN
30	181	2.0	1658	1	YM67_YEAST
31	180	2.0	3122	1	DPO2_MOUSE
32	178	2.0	2230	1	GOGA_HUMAN
33	178	2.0	2476	1	ATRX_MOUSE

RESULT: 1
KI67_HUMAN
ID KI67_HUMAN STANDARD; PRT; 3256 AA.
AC P46013;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Antigen KI-67.
GN MKI67.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94043435; PubMed=8227122;
RA Schluter C., Duchrow M., Wohlenberg C., Becker M.H.G., Key G., Flad H.-D., Gerdes J.;
RT "The cell proliferation-associated antigen of antibody Ki-67: a very large, ubiquitous nuclear protein with numerous repeated elements, representing a new kind of cell cycle-maintaining proteins.";
RL J. Cell Biol. 123:513-522(1993).
RN [2]
RP SEQUENCE OF 1-31 FROM N.A.
RA Gerdes J.;
RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: THOUGHT TO BE REQUIRED FOR MAINTAINING CELL PROLIFERATION.
CC -!- SUBCELLULAR LOCATION: NUCLEAR. PREDOMINANTLY LOCALIZED IN THE G1 PHASE IN THE PERINUCLEAR REGION, IN THE LATER PHASES IT IS ALSO DETECTED THROUGHOUT THE NUCLEAR INTERIOR, BEING PREDOMINANTLY LOCALIZED IN THE NUCLEAR MATRIX. IN MITOSIS, IT IS PRESENT ON ALL CHROMOSOMES.
CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -!- DEVELOPMENTAL STAGE: EXPRESSION OF THIS ANTIGEN OCCURS PREFERENTIALLY DURING LATE G1, S, G2, AND M PHASES OF THE CELL CYCLE, WHILE IN CELLS IN G0 PHASE THE ANTIGEN CANNOT BE DETECTED.
CC -!- SIMILARITY: CONTAINS 1 FHA DOMAIN.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC EMBL; X65550; CAA46519.1; -;
DR EMBL; X65551; CAA46520.1; -;
DR EMBL; X94762; CAA64388.1; -;
DR MIM; 176741; -;
DR InterPro; IPR000253; FHA_domain.
DR Pfam; PF00498; FHA; 1.
DR SMART; SM00240; FHA; 1.
DR PROSITE; PS50006; FHA_DOMAIN; 1.

34 177.5 2.0 1972 1 P531_HUMAN Q12888 homo sapien
35 176.5 1.9 3358 1 PGCV_MOUSE Q62059 mus musculu
36 176 1.9 2704 1 BPAL_HUMAN Q03001 homo sapien
37 175.5 1.9 2845 1 APC_MOUSE O61315 mus musculu
38 175 1.9 1210 1 AF4_HUMAN P51825 homo sapien
39 174.5 1.9 1902 1 SMF1_HUMAN O14497 homo sapien
40 174 1.9 728 1 TRDN_MOUSE Q13061 homo sapien
41 173.5 1.9 839 1 GLT5_WHEAT P10388 triticum ae
42 173.5 1.9 3210 1 CENF_HUMAN P49454 homo sapien
43 173 1.9 2116 1 MYS2_DICDI P08799 dictyosteli
44 173 1.9 2414 1 P300_HUMAN Q09472 homo sapien
45 173 1.9 3866 1 HRX_MOUSE P55200 mus musculu

ALIGNMENTS


```

Db 3038 RKSSEPVIMKRSYEL-PSHRLDMDKDPDPSPYLLALWTPGETA----- 1579
QY 1532 LKNISRLRTEHOVYEL-PSHRLDMDKDPDPSPYLLALWTPGETA----- 1579
Db 3093 SRQDKTEAQOITEVFLAEIRIENRNEKPKMKTSPEDIQNDPDGARKPIPRDKVTE 3151
QY 1580 -----NSAOPP-EQKCGGKASGKMCFD-----ETCSECNLSREANSQTVR 1618
Db 3152 NKRCLRSARQNESSQPKVAESGGKSAKVLNOKNGKGEAGNSDSMCLRSRKTSPQAA 3211
QY 1619 GTL 1621
Db 3212 STL 3214

RESULT 2
ATTRX_HUMAN
ID ATTRX_HUMAN STANDARD; PRT: 2492 AA.
AC P46100; P51068; Q15886; Q9NTS3; Q9H0Z1;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Transcriptional regulator ATTRX (X-linked helicase II) (X-linked
DE nuclear protein) (XNP) (Znf-HX).
GN ATTRX OR RAD54L OR XH2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS 1; 2; 3; 4 AND 5), AND VARIANT S-1860,
RP AND VARIANTS ATR-X.
RX MEDLINE=97123494; PubMed=8968741;
RA Picketts D.J., Higgs D.R., Bachoo S., Blake D.J., Quarrell O.W.J.,
RA Gibbons R.J.;
RT "ATTRX encodes a novel member of the SNF2 family of proteins: mutations
RT point to a common mechanism underlying the ATR-X syndrome.";
RL Hum. Mol. Genet. 5:1899-1907(1996).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORMS 2 AND 4).
RX MEDLINE=97386582; PubMed=9244431;
RA Villard L., Lossi A.-M., Cardoso C., Proud V., Chiaroni P.,
RA Collea L., Schwartz C., Fontes M.;
RT "Determination of the genomic structure of the XNP/ATTRX gene encoding
RT a potential zinc finger helicase.";
RL Genomics 43:149-155(1997).
RN [3]
RP SEQUENCE OF 860-2492 FROM N.A.
RX MEDLINE=9317911; PubMed=7874112;
RA Stayton C.L., Dabovic B., Gullisano M., Gecz J., Broccoli V.,
RA Giovanazzi S., Bossolasco M., Monaco L., Rastan S., Boncinelli E.,
RA Bianchi M.E., Consalez G.G.;
RT "Cloning and characterization of a new human Xq13 gene, encoding a
RT putative helicase.";
RL Hum. Mol. Genet. 3:1957-1964(1994).
RN [4]
RP PRELIMINARY PARTIAL SEQUENCE FROM N.A.
RX MEDLINE=94214473; PubMed=8162050;
RA Gecz J., Pollard H., Consalez G., Villard L., Stayton C.L.,
RA Millasseau P., Khrestchatskiy M., Fontes M.;
RT "Cloning and expression of the murine homologue of a putative human
RT X-linked nuclear protein gene closely linked to PGK1 in Xq13.3.";
RL Hum. Mol. Genet. 3:39-44(1994).
RN [5]
RP SEQUENCE OF 2401-2492 FROM N.A., AND VARIANTS ATR-X.
RX MEDLINE=95211835; PubMed=7697714;
RA Gibbons R.J., Picketts D.J., Villard L., Higgs D.R.;
RT "Mutations in a putative global transcriptional regulator cause X-
RT linked mental retardation with alpha-thalassemia (ATR-X syndrome).";
RL Cell 80:837-845(1995).
RN [6]
RP SEQUENCE OF 1375-2492 FROM N.A.

```

```

RA Pearce A., Chapman J.;
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
RN [7]
RX E2H2 BINDING.
RX MEDLINE=98167853; PubMed=9499421;
RA Cardoso C., Timsit S., Villard L., Khrestchatskiy M., Fontes M.,
RA Collea L.;
RT "Specific interaction between the XNP/ATR-X gene product and the SET
RT domain of the human E2H2 protein.";
RL Hum. Mol. Genet. 7:679-684(1998).
RN [8]
RP SUBCELLULAR LOCATION, AND ASSOCIATION WITH PERICENTROMERIC
RP HETEROCHROMATIN.
RX MEDLINE=20040663; PubMed=10570185;
RA McDowell T.L., Gibbons R.J., Sutherland H., O'Rourke D.M.,
RA Bickmore W.A., Pombo A., Turley H., Gatter K., Picketts D.J.,
RA Buckle V.J., Chapman L., Rhodes D., Higgs D.R.;
RT "Localization of a putative transcriptional regulator (ATTRX) at
RT pericentromeric heterochromatin and the short arms of acrocentric
RT chromosomes.";
RL Proc. Natl. Acad. Sci. U.S.A. 96:13983-13988(1999).
RN [9]
RP DISEASE.
RX MEDLINE=20213147; PubMed=10751095;
RA Villard L., Fontes M., Ades L.C., Gecz J.;
RT "Identification of a mutation in the XNP/ATR-X gene in a family
RT reported as Smith-Fineman-Myers syndrome.";
RL Am. J. Med. Genet. 91:83-85(2000).
RN [10]
RP VARIANT ATR-X SER-1713.
RX MEDLINE=97196774; PubMed=9043863;
RA Villard L., Lacombe D., Fontes M.;
RT "A point mutation in the XNP gene, associated with an ATR-X phenotype
RT without alpha-thalassemia.";
RL Eur. J. Hum. Genet. 4:316-320(1996).
RN [11]
RP VARIANT JM GLN-2131.
RX MEDLINE=96224392; PubMed=8630485;
RA Villard L., Gecz J., Mattei J.-F., Fontes M., Saugier-Verber P.,
RA Munnich A., Lyonnet S.;
RT "XNP mutation in a large family with Juberg-Marsidi syndrome.";
RL Nat. Genet. 12:359-360(1996).
RN [12]
RP VARIANTS ATR-X.
RX MEDLINE=97467722; PubMed=9326931;
RA Gibbons R.J., Bachoo S., Picketts D.J., Aftimos S., Asenbauer B.,
RA Bergoffen J., Berry S.A., Dahl N., Fryer A., Keppler K., Kurosawa K.,
RA Levin M.L., Masuno M., Nerl G., Pierpont M.E., Slaney S.F.,
RA Higgs D.R.;
RT "Mutations in transcriptional regulator ATRX establish the functional
RT significance of a PHD-like domain.";
RL Nat. Genet. 17:146-148(1997).
RN [13]
RP VARIANT ATR-X LEU-246.
RX MEDLINE=20123062; PubMed=10660327;
RA Fichera M., Romano C., Castiglia L., Failla P., Ruberto C., Amata S.,
RA Greco D., Cardoso C., Fontes M., Ragusa A.;
RT "New mutations in XNP/ATR-X gene: a further contribution to
RT genotype/phenotype relationship in ATR/X syndrome.";
RL Hum. Mutat. 12:214-214(1998).
RN [14]
RP VARIANT SHS LYS-1742.
RX MEDLINE=99347960; PubMed=10417298;
RA Lossi A.-M., Millan J.M., Villard L., Orellana C., Cardoso C.,
RA Prieto F., Fontes M., Martinez F.;
RT "Mutation of the XNP/ATR-X gene in a family with severe mental
RT retardation, spastic paraplegia and skewed pattern of X inactivation:
RT demonstration that the mutation is involved in the inactivation
RT bias.";
RL Am. J. Hum. Genet. 65:558-562(1999).
RN [15]
RP VARIANT CWS THR-2050.
RX MEDLINE=99326061; PubMed=10398237;

```


or send an email to license@isb-sib.ch).

CC EMBL; AB029290; BAA83821.1; -
CC EMBL; AF141968; AAF06360.1; -
DR EMBL; AB033077; BAA86565.1; -
DR EMBL; AL137853; CAC15920.1; -
DR EMBL; AB007934; BAA32310.1; -
DR HSP; Q01082; IBKR.
DR InterPro; IPR001589; Actinin_act_bind.
DR InterPro; IPR001715; Calponin_hom.
DR InterPro; IPR002048; EF-hand.
DR InterPro; IPR003108; GAS2.
DR InterPro; IPR001452; SH3.
DR InterPro; IPR002017; Spectrin.
DR Pfam; PF00307; CH; 2.
DR Pfam; PF00036; efhand; 2.
DR Pfam; PF02187; GAS2; 1.
DR Pfam; PF00435; spectrin; 35.
DR SMART; SM00033; CH; 2.
DR SMART; SM00054; Efh; 2.
DR SMART; SM00243; GAS2; 1.
DR SMART; SM00150; SPEC; 35.
DR PROSITE; PS00019; ACTININ_1; 1.
DR PROSITE; PS00020; ACTININ_2; FALSE_NEG.
DR PROSITE; PS50021; CH; 2.
DR PROSITE; PS00018; EF_HAND; 2.
DR PROSITE; PS50002; SH3; FALSE_NEG.
KW Actin-binding; Cytoskeleton; Calcium-binding; Repeat; SH3 domain.
FT DOMAIN 1 295
FT DOMAIN 78 181
FT DOMAIN 194 295
FT REPEAT 314 355
FT REPEAT 591 623
FT REPEAT 680 784
FT REPEAT 786 800
FT DOMAIN 871 923
FT REPEAT 1250 1272
FT REPEAT 1287 1342
FT REPEAT 1455 1534
FT REPEAT 1547 1659
FT REPEAT 1815 1891
FT REPEAT 1932 2042
FT REPEAT 2260 2280
FT REPEAT 2372 2395
FT REPEAT 2398 2507
FT REPEAT 2510 2618
FT REPEAT 2621 2728
FT REPEAT 2731 2838
FT REPEAT 2841 2945
FT REPEAT 2987 3024
FT REPEAT 3136 3163
FT REPEAT 3187 3274
FT REPEAT 3277 3383
FT REPEAT 3386 3492
FT REPEAT 3495 3601
FT REPEAT 3604 3673
FT REPEAT 3713 3819
FT REPEAT 3832 3927
FT REPEAT 3982 4043
FT REPEAT 4046 4152
FT REPEAT 4155 4262
FT REPEAT 4265 4371
FT REPEAT 4374 4481
FT REPEAT 4484 4590
FT REPEAT 4593 4700
FT REPEAT 4707 4808
FT REPEAT 4811 4917
FT REPEAT 4920 4985
FT REPEAT 5030 5054
FT REPEAT 5096 5107
FT CA_BIND 5132 5143
FT CA_BIND 5276 5283
FT DOMAIN 5355

4 X 4 AA TANDEM REPEATS OF [GS]-S-R-[AR].

FT	CONFLICT	1	72	MSSDETLSSRSRSRSCRSRSYRSRSGSLSPCPGPD
FT				TLPMNLPHGEQKRRKSDVLDPAERAVRV -> MFVPLW
FT				AGIPGRDVGSLQPLPPGPKQKCTASRAVAVI (IN REF.
FT				2).
FT	CONFLICT	575	594	VAISSEDEGNLRFVVELLS -> GPSAPLKMKAISDLCLMN
FT				YCL (IN REF. 1).
FT	CONFLICT	1487	1487	A -> T (IN REF. 2 AND 3).
FT	CONFLICT	1963	1963	V -> A (IN REF. 1).
FT	CONFLICT	2052	2052	E -> D (IN REF. 1).
FT	CONFLICT	2083	2083	E -> K (IN REF. 2).
FT	CONFLICT	2290	2290	M -> V (IN REF. 2).
FT	CONFLICT	2321	2321	C -> Y (IN REF. 1).
FT	CONFLICT	2341	2363	MISSING (IN REF. 2).
FT	CONFLICT	2344	2350	SILPSVG -> EYRLFKI (IN REF. 3).
FT	CONFLICT	2523	2523	Q -> R (IN REF. 2 AND 4).
FT	CONFLICT	4670	4670	S -> T (IN REF. 2).
FT	CONFLICT	4833	4833	MISSING (IN REF. 2).
SQ	SEQUENCE	5430	AA; 620346	MW; 91ADB7F7580B440B CRC64;

Query Match 2.4%; Score 219.5; DB 1; Length 5430;
Best Local Similarity 17.5%; Pred. No. 0.0041;
Matches 323; Conservative 296; Mismatches 702; Indels 527; Gaps 78;

QY 16 TEQNDGSRDVLLEFDLNTPOQKSKRRKKMPKV-----VVEGPKRKKPKPAELPKV 69
DB 1577 TQQDLSALQKNSDL-KDLQDDIQNRATSFATVVKDIEGFMEENOTKLSPRELTAL--- 1632
QY 70 VVEGPKRPRKAATQEKVKSETGSAKKKLNKESATKKPANVGDMNSKSPVTLKSCRK 129
DB 1633 -----REKLHQAEQYEAQOEETRAQKE-LEEAVT-----S 1663
QY 130 ALNFDLENPGDARQGDSESEIVQNSSGANSFSEIRDAIGGTNGSFDSFVSQIDKTNGLGA 189
DB 1664 ALQOETEKSKRAK-----ELAENKKIDALDWTSVGSSGQLLTNLPGMEQLSGASL 1717
QY 190 MNQPLEVSMG-----NQ-PDKLSTGAKLADQOPDLLTRNQOQCFPVATQNTQFPMENQQA 244
DB 1718 EKGALDITDVGNGVNAPEKLDKQCEMMKARHOELL--SQQNFILATQSAQAFLDQGH 1775
QY 245 WL-----OMKNOLIGFPFGNQOPRMTIRNQOCLANGNQOPMYLIGTRPALVS----- 293
DB 1776 NLTPEEQMLQOKLG-----ELKEQYSTSLAQSEAEKQVOTLQDELQKFLQDHK 1825
QY 294 -----GNQQLGGPOGQNKRPFLNHTQCLPAGNQLYSGPTDMHQLVMSTGGQQ 340
DB 1826 EPESWLERSEKLENNHKGSSPETLPSLLKQ-----GS---FSEDVSHKGD 1872
QY 341 HGLLKNQO-----PGSLIRGQPCV---PLIDQOQATPKGFT-----HLNQMVA 382
DB 1873 REVTTISGQKVLDMENSPKREGKEPSEIGNLVKDKLDATERYTALHSKCTRLGSHLMLIG 1932
QY 383 TSMSSPGLRPHSQSOVPTTYLHVESVSRILNCTTCTCORSRAPAYDSLOODHOGNKYIL 442
DB 1933 -----QYHQFQNSADSLQAMWQACEAN-----VEKLLSDTVASDPGVL 1970
QY 443 SHEISNGCKKALPQNSLPTPIMAK-----LEEARGKROYHRAMGOTEKHDLNLAQOI 498
DB 1971 QEOLATTQLOEELAEH-QVPVEKLOKVARDIMELEGEPADPHRHVQETTSILSHFQSL 2029
QY 499 AQSQDVERHNSSTVEYLDAAKTKIQV-----VQENLHGMPPVEVIEIDDPDGDGARK 552
DB 2030 S-----YSLAERSLLQRAIAQSQVQESLESLSLQSIGVEQN-----LE 2069
QY 553 GKNATASISKAGSKGNSSPVKKTAEKICIVPKTAKKGRAGKKSVPFPAHASETQLNQP 612
DB 2070 GKQVSSLSGVI-----QEALATNMKLDIARQKSSL--EATREMYTRFME 2114
QY 613 TPKPTPLSRSPKGGKRSIQDSGKARGPSGELLQODSTAEIYRMONLYGDKEREQEQ 672
DB 2115 TADSTAA-----VLOKLAEVSRFEQLCLOQK----- 2145
QY 673 NAMVLYKGDALVPYSEKRRKPRKVDIDDETT-RIWNLMLMGKGEKGEDEKKEK 731

Db	2146	-----ESSLKKLLFOAEMFHLGSKLQOFMKNKSRMLASQNDQDITHF	2190
Qy	732	WEERRVFRGRADSFIAHMLVQGDREFSPWKGVSVDVGVFTQNVSDHLSASFMSL	791
Db	2191	FOQ-----IOELNLEMDQOEN-----LDTL-----EHLVTELSGCFALD	2226
Qy	792	AARPPPKLSSREDERNVRSVVE---DPEGCILNLNE-----IPSWQEKVQ---HPS	838
Db	2227	LCQHQRVONLRKDFTELQTKVREKENDASCCQQLDFFERKLVRTFKWLKETEGSIPPT	2286
Qy	839	DMEVSGVD-----SGSKEQRLDSCNSG---TERFNFLEKSNLNEEVLSSQD-----SFD	886
Db	2287	ETMSAKELEKQIEHLKSLDDWASKTLVEEINCKGTSLENLIMEITAPDSQKGTGSIL	2346
Qy	887	PAIFQSCGRVG---SC-----SCSKSADAEFP-----TTRCET---KT	917
Db	2347	PSVGSVGSVNGYHTCKDLTIQICMDSDVNLKYEKLGVLHERGESLQAILNRMEEVHKE	2406
Qy	918	VSQTSQSVOT-----GSPNLDEICLOQNER-----PHLYEGSGDVQ---KQETT	960
Db	2407	ANSVLQWLESKEEVLKSDAMSSPTKETVKAQAESKAFLAELQNSPKIQVKKEALAG	2466
Qy	961	VAQKPPDLEKTMNWKDSVCFQPPRNDNW-----QTPSSSYEQC-----ATR	1003
Db	2467	LLVTYPNSQEAENKKI-----QEELNSRWERATEVTVARQRLSEASHLACFOAAESQL	2522
Qy	1004	QPHVLDIE-DFGMGEGLEGYSWMSISPRVDKVNKNVPRREFRGGSPVREFTQIIPST	1062
Db	2523	QPLMEKELMGVILGP-----LSIDPNMLNAQKQV---QWMLKEFFARRQHQHLENEA	2574
Qy	1063	PHPLPGMH-LSGSSSAVOEHODDTQHNOODEMNKASHLQKTLFDLLNSSECLTRQSGTK	1121
Db	2575	QGILTPGQVSLSTSQVKE-----LQSIQKWWELTDKLSNR	2612
Qy	1122	QNITDGLPRDRTADVDVPLSNSSL---QNILVESNNKEQTAVEYKETAAILREMK	1179
Db	2613	SSQIDQAVTKSYQVQELLQDLSEKVRVAVGRLSVQSAISTQPEAVKQOLETSEIRSDLE	2672
Qy	1180	GTADGKKPTQWDSLRKDVNEGROBRKNMDSIDYAIRRASISEISEAIKERGMN	1239
Db	2673	QLDHEVKEAQTLCDLSVLI-GEQYKDELKRLKLETV-----ALPLOGLE	2716
Qy	1240	NMLAVRIKDFLERIVKDHGGIDL-----EMLRESPPDKAKDYLLSIRGLGKS-----VE	1289
Db	2717	DLAADRLNRLQAALASTQOFOQMFDELRTLDDKQSQQAKNCPISAKLERLSQLOQNEE	2776
Qy	1290	CVRLTLHLNLAFFVDTNVRTAVRMGWVPLQPLPESLQHLHLELYPVLESIOKFLWPRLC	1349
Db	2777	FOKSLNQHSGSYEIVAEGESLLLS--VPPGEEKRTLQNLQVLELKNHWEELSKKTADRS	2834
Qy	1350	KL-----DORTLYELHYQLITFGKVCTKSRPNCNACPMRGCRHPASAYASARLALPAPE	1405
Db	2835	RLKDCMQAKQYQWHE-----DLVPWIEDCK-----AKMSRLVTLD---2872	
Qy	1406	ERSLTSATIPVPPEFPVPAIMIELPLPLEKS-LASGAPSN-----RNCPEITIEEPASP	1460
Db	2873	-----PVQLESLLRSKAMLNLEVEKRRSLLEILNSAADI	2906
Qy	1461	GOECTEIPESDIED--AYNEDPD---EIPIKLNIEFGMTLREHME--RNMELQ-EG	1511
Db	2907	LINSSEADEGHRDEKAGINQNMADVTELOAKTQGSLEMTQRLREFQESFKNIEKKVEG	2966
Qy	1512	DMSKALV--ALPHTTTSIPTPKLNISRLRTEHQVVEYELPDS-----HRLLDGMDKREP-	1563
Db	2967	AKHQLEIFDALQSQACS-----KNLEKLRAQOEVLQALEPOQVDYLRNFTQGLVEDAPDG	3021
Qy	1564	-DPSYLLAIWTPGETANSAPPEQKCGKAGKWCDEFCTSCNSLREANSQTVRGTLIL	1622
Db	3022	SDASQQL-----HQEAQAQOE-----FLEVQKRVNS-----GCVN	3051
Qy	1623	IFCRTAMRGSFPLNGTYFQVNFLEFADHESLKPIDVPRDWIWDLPRT	1670

Db	3052	MENKLEGIGOF-----HCRVREMF-----SQLADLDDDLGCMGAIGRDT	3090
RESULT	5		
APC_RAT			
ID	APC_RAT	STANDARD;	PRT; 2842 AA.
AC	P70478;		
DT	16-OCT-2001 (Rel. 40, Created)		
DT	16-OCT-2001 (Rel. 40, Last sequence update)		
DT	16-OCT-2001 (Rel. 40, Last annotation update)		
DE	Adenomatous polyposis coli protein (APC protein).		
GN	APC.		
OS	Rattus norvegicus (Rat).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.		
OX	NCBI_TaxID=101116;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=FISCHER 344/N; TISSUE=Brain;		
RX	MEDLINE=96116966; PubMed=8563176;		
RA	Toyota M., Ushijima T., Kakiuchi H., Watanabe M., Imai K., Yachi A.,		
RA	Sugimura T., Nagao M.;		
RT	*cDNA cloning of the rat APC gene and assignment to chromosome 18.;"		
RL	Mamm. Genome 6:746-748(1995).		
RN	[2]		
RP	MUTAGENESIS.		
RC	STRAIN=SPRAGUE-DAWLEY, AND FISCHER 344/N;		
RX	MEDLINE=95148647; PubMed=7846077;		
RA	Kakiuchi H., Watanabe M., Ushijima T., Toyota M., Imai K.,		
RA	Weisburger J.H., Sugimura T., Nagao M.;		
RT	*Specific 5'-GGA-3'-->5'-GGA-3' mutation of the APC gene in rat colon		
RT	tumors induced by 2-amino-1-methyl-6-phenylimidazo[4,5-b]pyridine.;"		
RL	Proc. Natl. Acad. Sci. U.S.A. 92:910-914(1995).		
CC	-1- FUNCTION: TUMOR SUPPRESSOR. ALLOWS THE RAPID TURNOVER OF BETA-		
CC	CATENIN. APC ACTIVITY CORRELATED WITH ITS PHOSPHORYLATION STATE		
CC	ALLOWS THE DOWN-REGULATION OF CYTOPLASMIC BETA-CATENIN (BY		
CC	SIMILARITY).		
CC	-1- SUBUNIT: FORMS HOMOLOGOMERS AND ASSOCIATES WITH CATENINS (BY		
CC	SIMILARITY).		
CC	-1- PTM: PHOSPHORYLATED BY GSK-3b (BY SIMILARITY).		
CC	-1- SIMILARITY: CONTAINS 7 ARM REPEATS.		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/		
CC	or send an email to license@isb-sib.ch).		
CC	-----		
DR	EMBL; D38629; BAA07609.1; -.		
DR	HSP; Q02248; 3BCT.		
DR	InterPro; IPR000225; Armadillo.		
DR	Pfam; PF00514; Armadillo_seg; 6.		
DR	SMART; SM00185; ARM; 5.		
DR	PROSITE; PS50176; ARM_REPEAT; 1.		
KW	Anti-oncogene; Phosphorylation; Coiled coil; Repeat.		
FT	DOMAIN 1 728		
FT	DOMAIN 1 62		
FT	DOMAIN 125 260		
FT	REPEAT 451 493		
FT	REPEAT 503 545		
FT	REPEAT 546 589		
FT	REPEAT 590 636		
FT	REPEAT 637 681		
FT	REPEAT 682 723		
FT	REPEAT 724 765		
FT	DOMAIN 739 2831		
FT	DOMAIN 1130 1155		
FT	DOMAIN 1556 1575		
FT	DOMAIN 1864 1891		
FT	MUTAGEN 523 523		
FT	SEQUENCE 2842 AA; 310530 MW; 3CBB2EA8A34E8F47 CRG64;		

[illegible]

Qy	897	GSCSCSKDAEFTPTTCETKTVTSGTSQSVQTSQSPNLSDBICLQGNNE-----RPHL	946
Dy	1788	RKNTDSKVNTYNTBETTSNDK--DSKQSLKNPNKDLNDK--LPNDNEVRVGFTFDSPHH	1843
Qy	947	Y---EG-----SGDVQKQETTNNVAOKKPDLEKTMNWKDSVCFQPRNDTNQOT	991
Db	1844	YAPTEGTPYCFSRNDSLSDFD-DDVDLSREKAEALRKGESKDS-----EAKVTCHT	1896
Qy	992	TPSSSYE-----QCATROPHVLVDIEDFGMQGBGLGYSNMSTSPRVDYRKNKNVRRFRQ	1046
Db	1897	EPSSSQSQAKQAQASTKHP-----VNRGSPKLLQEQPTFPQ--SSKDVPDR----	1941
Qy	1047	GGSVPRFETGQIIPSPHPELPGMLGSLGSSSAVQEHQDDTHQNOQDEMKNKASHLOKTFDL	1106
Db	1942	GAATDEKLQNFALENTP-----VCFSRNSSLSDVDQENNNNEETGPV-----	1986
Qy	1107	LNSSEECCLTQSQSTKQNIITDGCILPRDRTAEDVVVDPLSNSSQLNITVBESNNSKEQTAVE	1166
Db	1987	--RDAEPANAQQPGKQASGYAPKSFHYEDTPVCFSRNSSLSDSD-----	2036
Qy	1167	YKETNATILRE-MKGTLAGKKPT-----SQNDSL-----KDVGNNGRQ	1206
Db	2037	-----LLRECISAMPKKRPSRLKGEGEQSPRKVGSVLAEDLTLDLKDIOQRPSEH	2089
Qy	1207	ERNKNMDSIDYEAIRASTSEISEAIKERGNMMLAVRIKDFLERIVKHGGIDLEWL	1266
Db	2090	GLSPDS--ENFDKAIQEGANSIVSSLUHAAAAACLSQAQSSDSITLSKSGVSLGSPF	2148
Qy	1267	ESPPDKARDYLLSTIRGLGLKSVECBRLLLLHNLAFPVDTNVGRTAVRMGWVPLQPLPESL	1326
Db	2149	HLTPDQEKFPFTHKG-----	2164
Qy	1327	QLHLELLPVLESIQKFLPRLCKLDQRTLYELH-----YQITFGKVC	1371
Db	2165	-----PRILPGEKSTLEAKTIESENKGIKGKKVYKSLITGKI--	2203
Qy	1372	TKSRPNCN-ACPMPGECRHFASAYASARLALPAPERSLTSATIPVPPESPPPVAIPMIE	1430
Db	2204	---RSNEISQMKQPLQTNWPSISRGRTWIHPGVRRNSSSTSPVSKKG-PP-----	2252
Qy	1431	LPLPLEKSLASG--APSNRENCEPIIEEPASQCECTEITESDIEDAYYNEDP-----D	1482
Db	2253	LKTPASKSPSEGPVATTPSRGTKPAVKSELP-----ITRQTSHISGSKNGKPSRSGSRD	2306
Qy	1483	EIPIKILNIOFGMTLEHMERNELOQDMKALVALHPTTTSPTP-KLKNLSR	1537
Db	2307	STPS-----RPTQOPLSRPMQ-----SPGRNSISPGRNGISTPNKLSQLPR	2347
RESULT 6			
NEST_RAT	ID	NEST_RAT	STANDARD; PRT; 1805 AA.
AC	AD	P21263;	
DT	DT	01-MAY-1991 (Rel. 18, Created)	
DT	DT	01-MAY-1991 (Rel. 18, Last sequence update)	
DT	DT	16-OCT-2001 (Rel. 40, Last annotation update)	
DE	NESTIN.		
GN	NEST.		
OS	Rattus norvegicus (Rat).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.		
OX	NCBI_TaxID=10116;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=90150286; PubMed=1689217;		
RA	Lendahl U., Zimmerman L.B., McKay R.D.G.;		
RT	"CNS stem cells express a new class of intermediate filament		
RT	protein".;		
RT	Cell 60:585-595(1990).		
CC	-!- TISSUE SPECIFICITY: CNS STEM CELLS.		
CC	-!- DEVELOPMENTAL STAGE: UPON TERMINAL NEURAL DIFFERENTIATION, NESTIN		
CC	IS DOWN-REGULATED AND REPLACED BY NEUROFILAMENTS.		

CC CC -!- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC CC or send an email to license@isb-sib.ch).
CC CC -----
DR EMBL; M34384; AAA41685.1; -;
DR PIR; A34736; A34736.
DR InterPro; IPR001664; IF.
DR Pfam; PF00038; filament; 2.
DR PROSITE; PS00226; IF; 1.
KW Intermediate filament; Coiled coil; Neurone.
FT DOMAIN 1 HEAD.
FT DOMAIN 8 314 ROD.
FT DOMAIN 315 1805 TAIL.
FT DOMAIN 8 43 COIL 1A.
FT DOMAIN 44 55 LINKER 1.
FT DOMAIN 56 151 COIL 1B.
FT DOMAIN 152 174 LINKER 12.
FT DOMAIN 175 193 COIL 2A.
FT DOMAIN 194 196 LINKER 2.
FT DOMAIN 197 314 COIL 2B.
SQ SEQUENCE 1805 AA; 198744 MW; B40EE14717E0998D CRC64;

Query Match 2.3%; Score 210.5; DB 1; Length 1805;
Best Local Similarity 18.7%; Pred. No. 0.0027;
Matches 283; Conservative 216; Mismatches 564; Indels 447; Gaps 72;
QY 60 PRKPAELPKVVEGPKRKRKAATQEKVSK--ETCSAKKNLKNESATKKPANVGDMN 117
Db 159 PRRP-----PAPHRIPGAPVEDELARLGEVVRGAVRDYQERVAHVHESLSGQARE 210
QY 118 KSPVET--LKSCRKALFNLENPGDARGQDSEISVONSSGANSFSEIRDAI--GGTNGSF 174
Db 211 RLSQAVRGARECR-----LEVQQLADROSLQRERALEORLSGRW 251
QY 175 LDSVSDQDKTN--GLGANNQPLEVSMGNQPKLSTGAKLARDQDPDLTRNQOQFPVATQ 233
Db 252 QDRLEATDKFQALVEALEQEKQGLSQIAQILEGGQGLAH-----LKMSLSLVATY 303
QY 234 NTQPFMENQQAOLQMKNLGIFPGNQPRMTIRNQOPCLAMGNQOPMYLIGTPRPALVS 293
Db 304 RTLEAENSR--LQTPGRGQASGLFLDPKL-----KPNF--LGIPEDQYL--GSLVPALSP 354
QY 294 GNQQLGGPQGNKRPI--FLNHQTCPLPAGNOLYGSPTDMHOLVMSTGGQHGHLIKNQOPG 351
Db 355 TSFSPPLPNTLETPTVTAFLKTQEFLOARTPTLAS--TPIPPISEAP-----CPN 402
QY 352 SLIRGQPCVPLIDQOATP---KGFTHLQNMVATSMSSPG-----LR 391
Db 403 AEVRAEVLPLSLQLOTAPELPLWKATVPSSAILPELEPFGKQKQGFHPPDLTSLATNLN 462
QY 392 PHSOSQVPTTVLHVESVRIILGTTGQSRAPAYDSLOQDIHOGNKYI-----LSHEI 446
Db 463 PHH-----PT-----LEAKDGESSRVSII--FQEDGQIWELVEKEADIEVKV 505
QY 447 SNGNGCKKALPQNSSLPTPIMAKLEAGS-----KROYHRAMGQTEKHDLNLAQQAQSQ 502
Db 506 ENSSAQK---TQESGLDT-----EETQDSQGPLQKQETLKAIG--EEPLMSLKIQNYETA 554
QY 503 DVERHNST-----CVEYLDAAKTKTKQVVQENL-----HGMP--PEVIEDDPTD 548
Db 555 GKNCNSSTEGHGLTEGPEKEKQIPLKSLFEKNVSEKTLNGLGVPLSELLGKEDTRTE 614
QY 549 -----GARK-----GKNTASISKASKGN-----SSPVKKT 574
Db 615 DQELMSPKGTILKRFSSLGKESQEVVVRFSKGNLBSWTFATKESQHPGLGFGAEDQMLERL 674

QY 575 AEKEKCIVPKTPAKKG-----RAGRKKSVPPPAH-----ASEIQLWQPTPPK 616
Db 675 VEREDQSFPSPEEDQACRPLQKQNEQPLGYEEASQILERLIEKESQESLRSPEED 734
QY 617 TPLSRPKPKG-----GRKSTQDSKGARGSGELLCQDSIAEIIYRMQNLYLGDK--EREQE 671
Db 735 QEAGRSIQKENQPLGYEEAEDQMLER-----LIEKESQESLSPENRIGKPLERENQ 789
QY 672 ONAMVLYKGCALVPYESKKRKPVKVDIDETTRIWNLLMGKDEKGD--EEKDKKKEK 730
Db 790 KSLRYLEENQETFPLESRNQPLRSLEVEEEOIRVPLEKVSQDSLSGLSAEENQPLR 849
QY 731 WVEEE-----RVFRGRADSFIAARMHLVQGRRRFSPMKGSVVDSVIGVFL-----TQNVS 780
Db 850 YLEEDDCINKSLLEDKTHKSIGLSLEDNRGDSIIIP-----QESQTVSLRPPPEEDRIV 904
QY 781 DHLSSAFMSLAARFPKPLSSSRREDENV--RSVVVEDPBGCIILNLNEIPSWQKVKOHP 838
Db 905 NHLEKES-----QEFSSRSSEEEQVMERSLEGENHE-----SLSSVEKEDQMVE--S 949
QY 839 DMEVSGVDSG-----SKQLRDCNSNGIERFNFLEKSIQNL-----EE 876
Db 950 QLEKESQDSKSLSEDESQETFGPLEKENAESLSLAGQDEEQKLEQETQOTLRAVNEQ 1009
QY 877 EVLSSQDSFDP-----AIFQSCGRVSGCS--CSKSDAEFPTR--CETKTVSGTS-- 922
Db 1010 MAVSPPEKVDPELPKPLGNQDEIARSLGKENQESLSLKEKGIETVKSLETEIIEPLETA 1069
QY 923 -----QSVOTGSPNLSDEICLOGNERPHLYE-----GSGDVOKQETTNVAQKKPDLK 971
Db 1070 EEDLERKRSIDTOEPLWSTVARETVPEPPEPGLSGVDENRETUTLSLEKESQELSSL 1129
QY 972 MNW-----KDSVCFQGPNDTN-----WOTTPSSVYEQ----- 999
Db 1130 GKNVETRVEDSQOCLQVEEGLOEHOESLRREVQBELPSSGNOQRWEDVVEGKAVGQEA 1189
QY 1000 -CATRQPHVLIDIEDFGMQGEG-----LGYSWM--SISPRVDRVKNKNVPR 1041
Db 1190 PLATTGVGTEDKAEHLRGGOGEEAAAEGELLQDIVGEAWSLGSSEPKQRPAPALDN 1249
QY 1042 RFRQGSVPRFTGQIIPSTPHLPOMGLSGSSSAVOEHODD--TOHQODEMKNASHLQ 1100
Db 1250 ---LEGAL-----EVP---VAQSMPEYTERDEDRQAAGEQDSIEVTILGLE 1289
QY 1101 --KTFDLNLSBECLTQSSSTKQNTDGLPRDRTAEDVVDPLSNNSLQNLIVSNSS 1158
Db 1290 AARTGLEL-----EQEVVGLDEPRHFAREBAIPPSLGESEVK----- 1326
QY 1159 NKEQTAVEYKETNATILREMKGTADGKPKTSQWDSLRKQVGE--NEGROERNKNNMDSID 1217
Db 1327 -----AKIAQGLEG---PGKEP-----KEAGALDSGILELPKTSSEALE 1362
QY 1218 YEAIRRASISEIAKERGMNNMLAVRIKDFLERIVKDHGGIDLEWLRSPPDKARD-- 1275
Db 1363 CQHEE---SESMEGWEDEAS-----LE--TSDHEGSDAPQPR--PETEEDG 1405
QY 1276 YLLSIRGLKSLVECVRLTLHLNLAFPVDTNVGRIAVRMGWVPLQPLPESLQLHLLELYP 1335
Db 1406 AQAALTPAGKLLIE-----FCSPILPILTAH--ELQP 1435
QY 1336 VLESIQKFLW 1345
Db 1436 QAEGIQEAGW 1445
RESULT 7
ANK2_HUMAN STANDARD; PRT; 3924 AA.
AC Q01484; Q01485;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Ankyrin 2 (Brain ankyrin) (Ankyrin B) (Ankyrin, nonerythroid).


```
Db 2336 AQESTATSDTKALPLPEASVYKTDGTGKESKQGVIRSPQGLLEALSPRDSSEVLAVADD 2395
Qy 291 LVSGNQOLGGQGNKRPIFLNHOCTLPAG-----NQLVGSPTDMHQLVMSGGQOHGLLI 345
Db 2396 LA-----VSHKSLSPVLENSHKTPDSL----- 2422
Qy 346 KNOQPSLIHQPCVPLIDQOATPKGFTHLNQMVATSMSSPGLRPHSQSOVPTTLHV 405
Db 2423 ---EPSPL--KESPCRDSSLESSVEPK-----MKAGIFP-SHPPLPAVAKT 2463
Qy 406 ESV-----SRILNGTGTQCRARAYDSLOD--IHQNKYILSHEI--SNGNGCKKA 455
Db 2464 ELLTEVASVRSLRDLDPGSAED-----DSLEQTSMESSGKSPSPDTPSSSEVSVEY 2517
Qy 456 LPONSLPTPTIMALEA-----RGSROYHRAMGTEKHDL-----NLAQQ 497
Db 2518 TPKTVDVTPPAVHECAEDDSNGEKKRF-----TPEEMFKVWTKMKMDELEQE 2571
Qy 498 IAQSDV-----ERHNSST-----CVEYLDAAKTKIKYVQENLHGMPEVIEIEDDP 546
Db 2572 AKQRDYKKEPKQBESSSSDPDADCSVDVDEPKHTG---SGEDSGVPLVVTSESRKV 2627
Qy 547 TDGARKGNATASISKASKG-----NNSPVKTAKEKCIYVK- 584
Db 2628 SSSSESEPELAQLKKGADSGLLPEFVIRVQPPSLPLPSMSDSNPSPEYQFQP---VVSQ 2684
Qy 585 -----TPAKKGRAGRKKS-----VPPAHASEIQLWQTPPKTPLRSRKP---KGKR 629
Db 2685 YTFKNEDTQEPKSEEEKSESHLAEDRAHVSTEAEDRSYKLNDRDTPQKICDHGC 2744
Qy 630 KSIODSGAR-----GPSGELICQDSIAE--IYRQNYLVLGDKEREQONAWLYKG 680
Db 2745 EAMSPSSARPVSSGLOSPTG---DDVDEQPVYIKESIALQGTHEKDTGEEL----- 2794
Qy 681 DGAIVPYEKRRKRPKVDIIDE---TTRIWNLMKGDEKGD----- 721
Db 2795 -----DVSRASPOADCSSEFSSSSSLPHCLVSEKELDEDISATSSIQTEVTKT 2846
Qy 722 EEDKKKEKWEERRVFRGRADF---IARMHLVQGDRRFSPWKGSVDS----- 769
Db 2847 DETFENLPKDCPSQSSITTTOTDRFSDVDPVPSDLAENDEIYDPOITSPYENVPQSFPSS 2906
Qy 770 -----VIGVFLTONVSDHL-----SSSAFMSLAARF----- 795
Db 2907 EESKTQTDANHTTFSHSEVYSVTITSPVEDVYVVASSSGTVLKSNEFEGQDIKMESQL 2966
Qy 796 -----PPLKSSREDERNVRSVV--EDPEGICILNLEI--PSWOR-----KV 834
Db 2967 ESTLWEMQSDSVSSFPETMSGATTVVGEQISKVITTKTDVDSWSWSEIREDDDAFAFARV 3026
Qy 835 QHPSDMEVSG--VDSGSK-----EQ-----LRDCSNSGI-- 861
Db 3027 KE-BEQKIFGLMVRQSGTTPDTTPARTPEEGTPTSEQNPFPLFQEGKLFEMTRSGAID 3085
Qy 862 -----ERNFL-----EKIQNLEEVLSQDSFDPATFQSCGRVSGSCSKSDA--- 906
Db 3086 MTKRSYADESHFFQIGESREETLSEVDKEGATGADPLPLETSAESIALSESKEETVDDE 3145
Qy 907 -----EFPTRCTKTVSGTSQS-----VOTGS-PNLS- 933
Db 3146 ADLLPDSVSEVEEIPASDAQLNSQMGISASTETPTKEAVSVGTKDLPTVGTGDIPLPSG 3205
Qy 934 -DEICLOGNERP-----HLYEGSDVQKQETTNNVAQKKPDLKRTMMNKDSVC 979
Db 3206 VKQICSPDSSEPAVQVQLDFSLTRSVYSDRGD--DSPDSSPEQKSVIEIPTAPMENVP 3263
Qy 980 FQQRPN-----DTNQWTPSSSYEQCATRQPHVLIDFQMGOGELGYSWMSISPRVD 1032
Db 3264 FTESKSIPTVMTPTSTPAPSAEYESSVS-----EDF-----LSSVDEENKAD 3307
Qy 1033 RVKNKN-----VPRFRFQ-----GGSVPREFT--GQIIPSTPHELPGMGLSGSSSAVQE 1080
Db 3308 EAKPKSKLPVKVPLQRFVEQQLSDLTQSVQKTVAPQGDMAASIAPDNRSKSESASSLDSK 3367
```

```
Qy 1081 HQDDTOHNOODEMNKASHLOKTFDLNLSSEECLETR-----QSSTKONITDQCL 1129
Db 3368 TKCPVKTRSTYETETESRERAELEL--ESEAGTRPKILTSRLPVKSRSTSSCRGTS 3425
Qy 1130 PRDRTAEDVVDPLSN-----NSSLQNLIVSNSSNKEQTAVYKETNATILREMGKT 1181
Db 3426 PTRKSEHFFDLVYRNSIEFFEFEISDEASKLVDRLTQSEREQEIVSDSSSA-----LEVS 3481
Qy 1182 LADCKKPTSDWSLRKDVNEGNEGRQERKNKNDMSIDYEAIRRASISISISAI-KERGMMN 1240
Db 3482 VIENLPPVETEHSPVEDI-----FDRPIWDESIEITLIERIPDENGHDH 3525
Qy 1241 MLAVRIKDFLERI-----VKDHGGI-----DLEWLRESPPDKAKD---YL 1277
Db 3526 --AEDPODEQERTEERLAYTADHLGFSWTELAELDFTEEQIHOIRIENPNSLODQOYL 3583
Qy 1278 LSI-----RGLGKSVCEVRLTLHLNLAFFVDVTNVIAGVAVRMGWVPLQLPESLQLHL 1330
Db 3584 LKIWLERDGGKHATDTNLVECLTKINRMDIVHLMETNT-----EPLQERISHSY 3631
Qy 1331 LELYP--VLSIQKF--LWPRCLKLDORTLYELHYQLITEGKVECTKSRPNCNACPMRGE 1386
Db 3632 AEIEQTITLDHSEGFSLQBELCTAQHKQKE--QAV-----SKESETCDHPPIVSE 3681
Qy 1387 CRHFASAYASARALAP-----APEERSL----- 1409
Db 3682 -EDISVGYSTFGQGVKPTGEGSSSTALFPQTHKEOVQDFSGKQMDLPESSLEYQOEYF 3740
Qy 1410 -----TSAT-----IPVPPESFPVVAIPMIELPLPLEKSLASGAPSNRENCEPIIEEP 1457
Db 3741 VTPGTSETQKAMIVPSPSKTP---BEVSTPAEEKLYLQTPTSSERGSGPIIQEP 3796
Qy 1458 ASPOECTE-----ITESD-----IEDAVY--NEDPDEIP---TIKLNIE 1492
Db 3797 EEPSEHREESPRKTSUVIVESADNQETCERLDEDAAFKGDMDPEIPPETVTEEYID 3856
Qy 1493 QFGMTLREHMERNM 1506
Db 3857 EHGHVTVVKVTRKI 3870
```

RESULT 8
AIM1_HUMAN STANDARD; PRT; 1723 AA.
ID AC Q9Y4K1; O00296;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Absent in melanoma l protein.
GN AIM1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=97250519; PubMed=9096375;
RA Ray M.E., Wistow G., Su Y.A., Meitner P.S., Trent J.M.;
RT "AIM1, a novel non-lens member of the betagamma-crystallin
RT superfamily, is associated with the control of tumorigenicity in human
RT malignant melanoma.";
RL Proc. Natl. Acad. Sci. U.S.A. 94:3229-3234(1997).
CC -I- FUNCTION: MAY FUNCTION AS SUPPRESSOR OF MALIGNANT MELANOMA. IT MAY
CC EXERT ITS EFFECTS THROUGH INTERACTIONS WITH THE CYTOSKELETON.
CC -I- SIMILARITY: BELONGS TO THE BETA/GAMMA-CRYSTALLIN FAMILY. CONTAINS
CC 6 TANDEM REPEATS OF A BETA/GAMMA-TYPE CRYSTALLIN DOMAIN.
CC -I- SIMILARITY: CONTAINS 1 RICIN B-TYPE LECTIN DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation on
CC the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR ENBL; U83116; AAB3792.1; -;
DR ENBL; U83115; AAB3791.1; -;
DR HSSP; P02526; 1GAM.
DR MIM; 601797; -;
DR InterPro; IPR001064; Crystallin.
DR InterPro; IPR000772; Ricin_B_lectin.
DR Pfam; PF00030; crystall; 5.
DR Pfam; PF00652; Ricin_B_lectin; 5.
DR PRINTS; PR01367; BGCYSTALLIN.
DR SMART; SM00458; RICIN; 1.
DR SMART; SM00247; XTALBq; 6.
DR PROSITE; PS00225; CRYSTALLIN_BETAGAMMA; 7.
DR PROSITE; PS00231; RICIN_B_LECTIN; 1.
KW Repeat; Lectin.
FT DOMAIN 1022 1119 CRYSTALLIN BETA/GAMMA-LIKE 1.
FT DOMAIN 1123 1206 CRYSTALLIN BETA/GAMMA-LIKE 2.
FT DOMAIN 1207 1318 CRYSTALLIN BETA/GAMMA-LIKE 3.
FT DOMAIN 1319 1404 CRYSTALLIN BETA/GAMMA-LIKE 4.
FT DOMAIN 1405 1497 CRYSTALLIN BETA/GAMMA-LIKE 5.
FT DOMAIN 1502 1584 CRYSTALLIN BETA/GAMMA-LIKE 6.
FT DOMAIN 1586 1719 RICIN B-TYPE LECTIN.
SQ SEQUENCE 1723 AA; 188646 MW; 7550F681A627FB09 CRC64;

Query Match 2.3%; Score 207; DB 1; Length 1723;
Best Local Similarity 18.7%; Pred. No. 0.0038;
Matches 235; Conservative 170; Mismatches 475; Indels 376; Gaps 53;

QY 18 QNDSRODVLFEFLNKTPOQPKSRKPKMPKVVVEGPKRKPAPLKVVEGPKR 77
DB 91 ESDRSKO-----PPASPTKRKRGRSALRAVAPPASGPRAPAK-----ESPKR 136

QY 78 KPRKAATQEKVKSETGSAKKNLKESATKPPANVGDMNKSPEVTLKSKRKAL--NFDL 135
DB 137 VPDPSPVTKTAASGEAARAIPRELPAK-----SSILPEIKPKRGLPNHFN- 188

QY 136 ENPDARQGDSESEIVONSSGANSFSEIRDAIGTNGSFLDSVQIDKTNGLGAMNOPLE 195
DB 189 ---GRAEGGRSR-----ELGRAAGAPGASDADGLKPRNHF-GVGRSTVTTK 230

QY 196 VSMGNQPKLSTGAKLARDQDPLLTRNQCQEPVATQTFPM---ENQAWLQMKNL 252
DB 231 VTLPAKPKHVELNLTTPKNLD-SLGNHNPFSQPVHKGNTATKLSLFENKRT----- 281

QY 253 IGFPFGNQPRMT-IRNQPCLAMGNQOPMYLIGTPRALVSGNQOLGGPQGNKRP---- 307
DB 282 -----NSSPRHTDIRPNTPASSKT-----FVGRKLNLAKEAKEQEPKVMPSNQ 331

QY 308 -----IFLNHQTCLPAGNOLYSGPTDMHQL-----VNSTGG----- 338
DB 332 NGVLVKETAETKTVTSEETLTPATRMNGDSSENQALGPQPNODDKADVQTDAGCLSEP 391

QY 339 -----OOHGLLKNQPCSLIRGQPCVPLIDQDQATPKGFTHLNQVATSMSSPGL 390
DB 392 VASALIPVKHKLKEDESRRA-----DSKSLVLENTVDTAQDIPITVDTKDL 439

QY 391 RPHSQSQVPTTYLHVESVSRLNCTT-----GTCQRSRAPAYDSLQODTHQG 437
DB 440 PPTAMPKQHTFSDQSPASSPSPSLSAPAPGDVPKTCVQSPSSPCTDLKYSFN 499

QY 438 NKYLHSINGCKKALPONSLSPTTMA--KLEPARGSK-----ROYHRAMQG 486
DB 500 HKGCVL-PVSRQNNKMPLELGGTTPPLSTERSPEAVSECPVSRVLVQVRSVLPVES 558

QY 487 TEKHDNLAAQIAQSDOVERHNSSTC-----VEYLDAAKTKIKVVOENLHG----- 534
DB 559 TQ---DVS-SQVIPESSEVREVLPTCHNSEPEVVSASCAPPOVEVLGNHSHCTAEALAA 615

QY 535 -MPPEVIE-----IEDDPTGARGKGNKA----- 557
DB 616 KSGPQVTPASEKTLPTQAQSGSRTPLAESSTPNSPSSGNHLATPQRDQTVTNGQDS 675

QY 558 -----SISKASKG---NSSPVKKTAE-----KEKCIVPKTPAKKGRAGRK 595
DB 676 PASLLNISAGSDSDVDFSSDMKEFTETIKQMSAVCMVPMKRRKARMPNSPA-----P 728

QY 596 KSVPPPAHASEIQLWQTPPKTPLSRSKPK-----GKRKSIQDSGKARGSGELLQDSI 651
DB 729 HFAMPPHEDHLE-----KVDFPKVFTFGLGKK--KES-----QPEM 763

QY 652 AEIYRMQNLVLDGKEREQDNA--MVLYKGDGALVPYESSKKRPRKPKVDIDDETTRWN 709
DB 764 SPALHLMQNLDTKSLRPRKASAEQSVLEK---SLHTNTNGNSEPLVMPMEINDKENR--- 817

QY 710 LLMGKGDEKDEKDKKKEKWEERVRGRADSFIARMHLVQGDRRRSPKWSVDS 769
DB 818 -----DVTNGGIKRSRLE-----KSALFSSLLSLPODKIFSPSVTSVNTM 858

QY 770 VIGVFLTONVSDHLSAFAFMSLAARPPKLLSSREDENRNVSVVVEDPEGCILNLNEIPS 829
DB 859 TTAFTSTQNGS---LSQSVSQPTTEGAPCGLNKE-----QSNLLPDNSLKVFNFSST 911

QY 830 WQEKVQHPSDMEYVSDGSGKEQLRDCNSGIERFNFLEKSQNLNLEEVLSQSDSPDAI 889
DB 912 SHSLSKSPSHMEKVPQKEKTKEDLDSRSLHLPETKFE--LSKLKNDMEKANHIESVI 969

QY 890 FQSGRRVSGSCSKSDAE---FPTTRCEKTVSGTSQSQTGSPNLSDICLOGNERPH 945
DB 970 -----KSNLPCNANSDFDFMGLFKSSRYD-----PSIFSGLSDTMTLRGVSQNK 1016

QY 946 LYEGSGDVQKQETTNVAOKKPDLEKTMNWKDSVCFQGPDRDNTNQTTPSSSYEQCATRQ 1005
DB 1017 LNRPQKV-----VIYSEPDVS-----EKCI---E 1038

QY 1006 HVLIDIEFGMGEGGLGYSWMSISIPRDRVKNKNVPRFRFGGSGVPREFTCQIIPSTPHE 1065
DB 1039 VFSDIQDCS-----SW-SLSP-VILIKVVRGWCWILYEQ-----DNFEGHSIPLEGE 1083

QY 1066 LPMGLSGSSSAVQEHQD-----DTQHQODEMKNKASHLOKTFLLDLSNSE 1111
DB 1084 LELSLGLWIEDILEREHEAESDKPVVIGSIRHVQD--YRVSH-----IDLTFEPE 1132

RESULT 9
ZEP1_HUMAN STANDARD; PRT; 2717 AA.
AC P15822;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Zinc finger protein 40 (Human immunodeficiency virus type I enhancer-binding protein 1) (HIV-Ep1) (Major histocompatibility complex binding protein 1) (MBP-1) (Positive regulatory domain II binding factor 1) (PRDII-BF1).
DE (PRDII-BF1).
GN HIVEP1 OR ZNF40.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90169514; PubMed=2106471;
RA Fan C.M., Maniatis T.;
RT "A DNA-binding protein containing two widely separated zinc finger motifs that recognize the same DNA sequence.";
RL Genes Dev. 4:29-42(1990).
RN [2]
RP STRUCTURE BY NMR OF 2113-2142.
RX MEDLINE=91064333; PubMed=2248949;
RA Omichinski J.G., Clore G.M., Appella E., Sakaguchi K., Gronenborn A.M.;

KW DNA-binding; DNA repair; Nuclear protein; Zinc-finger; Polymorphism.	
FT ZN_FING 3042 3057	C4-TYPE (POTENTIAL).
FT ZN_FING 3086 3104	C4-TYPE (POTENTIAL).
FT VARIANT 231	Q -> H.
FT VARIANT 389	/FTID=VAR_008516.
FT VARIANT 1540	S -> T.
FT VARIANT 2607	/FTID=VAR_008517.
FT VARIANT 2607	K -> E.
FT VARIANT 2607	/FTID=VAR_008518.
FT VARIANT 2607	S -> T.
FT CONFLICT 237	/FTID=VAR_008519.
FT CONFLICT 1156	E -> Q (IN REF. 4 AND 5).
FT CONFLICT 3130	Y -> C (IN REF. 4 AND 5).
SQ SEQUENCE 3130	AA; 352782 MW; 1C0700900F10BB14 CRC64;
Query Match 2.2%; Score 200.5; DB 1; Length 3130;	
Best Local Similarity 18.3%; Pred. No. 0.019;	
Matches 345; Conservative 229; Mismatches 647; Indels 663; Gaps 88;	
QY 6	DSSAVNATEAQEQSDQVLEFLNKTQQKPSKRKRKEMPKVYVEGPKRKRPR--- 61
DB 511	DNLSLLASLSLPQLDGTADENSDNPLNENSRTHS-----SVIATSKLSVKPSIFHK 562
QY 62	-----KPAELPKVYVEGPKRKRKPKAATQEKVKSKE-----TGSAKKNLKSAT-- 106
DB 563	DAATLEPSSAKITFOCK-----HTSALSHVLNKIEDLSQTNKTEKGLDNSYTSF 617
QY 107	-----KPPANVGDM-----SNKSPVTLKCRKALNFLEN--PGDARQDGS- 146
DB 618	TNESTYSMKYPGSLUSTVHSENSHKKKEILPVSCESI--FDYEDIPSVTRQVPSR 676
QY 147	-----ESEIV-----QNSGANSFSEIRDAIGCTNGSFLDSVQIDKTNGLGA 189
DB 677	KYTNIRKTEKSDPFLMHRHPNENTLGNPF-----NFDLNSHKNKVS----- 720
QY 190	MNQPLEVSMGNQDKLSTGAKLARDQPDLLTRNQOQCFPVATONTOPFPHENQQAWLQMK 249
DB 721	-----SEGNEKNSTALSFL-----PSSFTENCELSCSGENRTMVHLSNSTADESGL 769
QY 250	NOLIGFPFGNOOPRTIRNQOPLAMGNQOPMYLIGTPRALVSGNOQLGGPQ----- 302
DB 770	NKL-KIRYEEQEHKT---EKPSLSQQAHYMFF-----PSVVLN-CLTRPQKLSPTY 819
QY 303	-----GNKRPIFLNHQTLCPAGNQLYSGPTDMHQLVMSTGGQGHLLIKNOOPGSLIRGQQ 358
DB 820	KLQPGNK-PSRLKLNKRLKLAGHQTSTKSS-----ETGSTKDNFIQNN----- 861
QY 359	PCVPLIDQOAPTPKGFTHLQNMVATSMSSPGLRPHSQSOVPTT-----YLHVESV 408
DB 862	PCNSNPEKDNALASDLTKTT-----RGAFENKTPTDGFDHFGDGTLETEQ- 908
QY 409	SRLNGTTGTQCRSRAPAY---DSLQODIHQGNKYILSHEISNGNG-----CKKALPON 459
DB 909	SFGLVGNKYTLARAKVNYETEDSESSFTVNSKISLPHMEIGESLDGTLKSRKRKMS 968
QY 460	SSLPTPIAK---LEEARGSKROYHRAMGOTEKHD-----LNLAQQIAQSOV--- 504
DB 969	KKLP-PVIKVIIVIIINRRPGRKNMLVK-LGKIDSKEQVILTEEKMELYKKLAPLKDFWPK 1026
QY 505	-----ERHNSSTCEVYLDRAKKT-KLQKVVOENLHGMPEVIEIDDP 546
DB 1027	VPDSPATKIPIYPLTPKSHRKS--KHKSAKKTKGQOORTNENI----- 1070
QY 547	TGARKGNTAISKASKGNSNPKVKT--AEKEKCIVPKTPAKKGRAGRKSVPPPAHA 604
DB 1071	-----KRTLSFRKRSKSHAILSPSPSYNAETEDCDLNYSDVMSKLGFLSERSTSPINS 1123
QY 605	SEIQLWQTPPKTPLRSKPKGKRKSIQDSKGKARGPSGELLQCDSDIAEITYRMONLY-- 662
DB 1124	SPRCWSPTDPR-----AEEIMAAAEKAMLFKGNVYKK 1158
QY 663	-----LGDKEREQONAMVLYGDGALV--PYESKKRPRPKVIDDETTRINWLLMGKG 715

DB 1159	TVNSRIGKTSRAAQ-----IKSKAKLANPSIVTKRNRK-----NOTNKLVD----- 1202
QY 716	DEKEGDEKDKKKEKWKWEEERRVFRGRADSFITARMHLVOGDRFRSPWKGVSVDVSLGVFL 775
DB 1203	DGKKKPRAKOKTNEKG-----TSRKHITLTKDEKIKSQSGAEVAFVL---K 1244
QY 776	TQNVSDHLSSAFMSLAARFPPKLSRSSREDNRVSVVVEDPEGCILNLNLEIPSWQEKVQ 835
DB 1245	HQNVSEFASSSGSQL-----LFQKQDPLMGSADV 1275
QY 836	HPSDMEV-SGVDSGSKQLRDCNSGIERENFLEKSIQNLEEEVLSSQDSFDPALFQSCG 894
DB 1276	HPLSASLPTGIN--AQOKLSGCFSSLES---KKSVD--LQTFPSSRDLHPSV--CN 1325
QY 895	RVGSCSCSKSDAEFPTTRC-----ETKTVSGTSSQSVQT 927
DB 1326	SIGP-GVSKINVQRPHNQSAMFTLKESTLIQKNIFDLNSHLSQVAQNTQISSGSSKIED 1384
QY 928	GPNLSDEICLOGNERPHLYEGSDVQKQETTVNAQKKPDKLEKTMW---KDS---VCF 980
DB 1385	NANNI-----QRNYL---SSIGKLSYRSLSKLDQAYTPNHLCKDSQQQIVCI 1432
QY 981	GQ-----PRNDTN-----WQTPSSSYEQCATRQPH---VLDIEDFGMQEG 1019
DB 1433	AEQSKHSETCSPGNATASEESQMPNCFVTSLRSPINQIAWEQKQRGFILDMSNFKPE--- 1489
QY 1020	LGYSWMSISPR-----VDRVKNKVVPRFRFGGSGVPREFTGQIIPSTPHELPG 1068
DB 1490	-----RVKPRSLSEALSQTKALSOCKNRNV-----STPSAF-----G 1521
QY 1069	MGLSGSSAVQEHQDDTQHNQODEMKNASHLQKTLFDLLANSSECLTROSSTKQNTIDGC 1128
DB 1522	EGOSG-----LAVLKELLOKROOKAON----- 1543
QY 1129	LPRDRTAEDVVDPLSN---NSSLQNLIVESNSNKBQTAVEYKETNATILREMKGTLD 1184
DB 1544	-----ANTQDPLSNKHQPNKNSGSL-EHNKANKRTRSV-----TS 1579
QY 1185	GKPTQOWDSLKRDVEGNEGRQERNKNMDSIDYEATRASISISIAIKERGMNNMLAV 1244
DB 1580	PRKPTPRSTKQKE-----KIPKLLKVDSLN---L 1606
QY 1245	RKDFLERIVKHGIDLEWLRESPPDKADYLLSIRGLGLKSECVRLTLHLNLAFPVD 1304
DB 1607	QNSQLDNSVSDSPFPF-----SDPGFESCYSLE-----DSLSPENHYNEDIN 1650
QY 1305	T--NVGRIAVRMG--WVPL-QPLPES-LQLHLELYPVLESIOK--FL----- 1344
DB 1651	TIGQTGFCFSYSGSQFVPADQNLQKFLSDAVQDLFPG-QAIEKNEFLSHDNQKCEDKH 1709
QY 1345	-----WPLRCKLD-----QRTLYELHYQLITFGKVCTKSRPN 1377
DB 1710	HTTDSASWIRSGTSLSPSEIFKSTIDSNEHRRHONKNSFHPLTTRSNSIMDSFCVQQAED 1769
QY 1378	CNACPMRGECHRFASASARLALPAPEE-----RSLTSAT-----IPVPP 1418
DB 1770	CLSEKSR---LNRSSVKEVFLSLPQPNNSDWIOGHTRKEMGQSLDSANTSFATLLSPD 1826
QY 1419	ESFPVPAIPMTEL-----PLPLEKSLASGAPNSREN-----CEPIIEPA 1458
DB 1827	GELVDVACEDLELYSVRRNDMLTTPDSSPRSTSSPSQSKNGSFTPTRTANILKPLM-SPP 1885
QY 1459	SPGECTEITSDIEDAYNE---DPDEIPTIKLNTIQFGMTLREHMERNEWLEQDMS 1514
DB 1886	SREEIMATLLDHLSETIYQEPFCNSPDSVPEKPREIGGRLLMVETRLANDLAEFEGDFS 1945
QY 1515	KALVALHPTTTSITPPLKNISRLRTEHQVVELPDSHRLLDGMDKRE-----PDDPSPYL 1569
DB 1946	LEGURLMKATFSAMTQNPDPGSPURSGOGVNVKNGSSNPKMWVEDKVIIMPCCKAPSRQL 2005
QY 1570	LAIWTPG--ETANSQAQPEQKCGG 1591

Db 492 PEGVSEVMSLQOERKMGVSPGLKLFSTGLKLGKQKGRGGDESGEHTQVPA 551
Qy 616 KTLPSRSGKGRKSTODSGKARPSGELL- QDSIAEIIYRMQNLGLDGEREQBNA 674
Db 552 DSPDSQEQKGSASSPEPE-----EITCLEKGLAEV-----QDGEAEBA 595
Qy 675 MVLYKGDG-----ALVPYSEKRRPKVDIDDTTRINLLMKGDEKDEKDKKKE 729
Db 596 ---TSDGKKREGVTPWASFKKMVTPKRV-----RRPSEKDELDKVK- 638
Qy 730 KWEEERVRFRGRADSTFIARHLVQDRRFPWKGVSVDVIGVFLTQNVSDHLSAFA 789
Db 639 -----SATLSSTP----- 646
Qy 790 SLAAREFPKLSRSDERNRVSVVDEPEGCILNINIPSMQEKV-----QHPS 838
Db 647 STASEMEEMKGSVEEPK-----PEPKRVDTVSWEALICVSGSKRARRRSS 697
Qy 839 DMEYSGVDSGSKQRLRDCNSGIERFNFLKSIQNLBEEVLSSODSFDPAIFOSCGRVGS 898
Db 698 DEEGPKAMGDHQKAD-EAG-----KDKETGTDGILAGSQEHDP-----GQ 738
Qy 899 CSCSKSDAEPTTRCETKTVSGTSQSVQTSQSPNLSDCLQGNRPHLYEGSGDVQKQET 958
Db 739 GSSPEQAGSPT-----EGEGSVTWSFRLVTPRKKSLEKSEDSI-AGSG----- 787
Qy 959 TNVAKKPDLEKTMNWKDSVCFGQPRNDTNW-----QTPSSSYEQCATRQPHV 1007
Db 788 --VEHSPTDTE-----PGKEESVSIKKFIPGRKKRPDKQEQA----- 825
Qy 1008 LDIEDFGQGGELGYSM-SISP--RVDVRKNKNVPRFRFGGGSVPR-----EFTQII 1059
Db 826 -PVEDAGTGANEDSDVPVAVPLSEYDAVEREKNEAQQAQKGAQEQPEQKAATEVSKELS 884
Qy 1060 PSTHELPGMGLSGSSA--VOEHQDDTQHQQDEMKNASHLQKTFDLLNSSECLTRQ 1117
Db 885 EQVHMAAAVADGTRATITIEERSPW-----ISASVTEPLEQV 924
Qy 1118 SSTQNTDGLPRDRAED-----VVDPLSNSSLQ--NILVESNKNKEQTAVEYKETN 1171
Db 925 EAAALTEEVLEVEVIAEPEPTVTEPLPENREARGDTVVSEAEALTPEAVTAAE---T 980
Qy 1172 ATILREMGTLADGKKPTSQWDSL-----RKDVEGNEGROERKNKNMDS 1215
Db 981 APLGSEEGTASAAEETEMVSAVSQLTSDPTTEATPVOEVEGGVDPDIEQERRTQE 1040
Qy 1216 IDYEAIRRASISSEATKER-----GMMNMLAVRKDLERIVKDHGIDLEWLRSPPD 1271
Db 1041 V-----LQVNAEKVESQLPGTGP-----EDVLQPVQRAEA-----ERPEE 1078
Qy 1272 KAKDYLLSIRGLKGSVECVRLTLHNLAFVVDNNGRI-AVRMGWVPLQPLPESLQHL 1330
Db 1079 QAE-----ASGLKKE-----DVVLKVDQAQEAKEPTQGVVQQTTPSEF--- 1120
Qy 1331 LELYPVLESIOKFLWPLCKLDQRTLYELHYQLITFGVKTRKSRPNACPMRGECHRF 1390
Db 1121 -KAPQVTESIES-----SELVT-----TCQAEFLAG----- 1145
Qy 1391 ASAYASARLALPAPERSITSATIPVPPSPPPVPAIPMIELPLP-----LEKS 1438
Db 1146 ---VKSQEMV-----EQAIIPDSVETPTDSDTSDTGVADFAQCTOKDELVEIHEENE 1198
Qy 1439 LASGAPSNRENCPI---IEEPASPG-----QECTEITESIEDAYNDDPD-EIPTIKL 1489
Db 1199 VASGTQSGTGAEAAPQAOKERPAPPSSEVFQEBTK-EQSKMEDTLEHTDKVSVETVSI 1256
RESULT 12
BRC1_MOUSE
ID BRC1_MOUSE STANDARD; PRT; 1812 AA.
AC P48754; Q60957; Q60983;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Breast cancer type 1 susceptibility protein homolog.
GN BRCAL.
OC Mus musculus (Mouse).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Embryo;
RX MEDLINE=96177659; PubMed=8634697;
RA Abel K.J., Xy J., Yin G.Y., Lyons R.H., Meisler M.H., Weber B.L.;
RT "Mouse Brcal: localization sequence analysis and identification of
evolutionarily conserved domains.";
RL Hum. Mol. Genet. 4:2265-2273(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6;
RX MEDLINE=96177660; PubMed=8634698;
RA Sharan S.K., Wims M., Bradley A.;
RT "Murine Brcal: sequence and significance for human missense
mutations.";
RL Hum. Mol. Genet. 4:2275-2278(1995).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=129/SVJ;
RX MEDLINE=96121367; PubMed=8575748;
RA Bennett L.M., Haugen-Strano A., Cochran C., Brownlee H.A.,
RA Fiedorek F.T. Jr., Wiseman R.W.;
RT "Isolation of the mouse homologue of BRCAL and genetic mapping to
mouse chromosome 11.";
RL Genomics 29:576-581(1995).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=129/SVJ; TISSUE=Embryo;
RX MEDLINE=96067162; PubMed=7590247;
RA Lane T.F., Deng C., Elson A., Lyu M.S., Kozak C.A., Leder P.;
RT "Expression of Brcal is associated with terminal differentiation of
ectodermally and mesodermally derived tissues in mice.";
RL Genes Dev. 9:2712-2722(1995).
RN [5]
RP SEQUENCE OF 727-1111 FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Embryo;
RX MEDLINE=96021028; PubMed=7550308;
RA Marquis S.T., Rajan J.V., Wynshaw-Boris A., Xu J., Yin G.Y.,
RA Abel K.J., Weber B.L., Chodosh L.A.;
RT "The developmental pattern of Brcal expression implies a role in
differentiation of the breast and other tissues.";
RL Nat. Genet. 11:17-26(1995).
RN [6]
RP SEQUENCE OF 789-1250 FROM N.A.
RC STRAIN=129/SVJ;
RX MEDLINE=96163506; PubMed=8566965;
RA Schroeck E., Badger P., Larson D., Wynshaw-Boris A.,
RA Ried T., Brody L.;
RT "The murine homolog of the human breast and ovarian cancer
susceptibility gene Brcal maps to mouse chromosome 11D.";
RL Hum. Genet. 97:256-259(1996).
CC -!- FUNCTION: NOT KNOWN, MAY REGULATE GENE EXPRESSION. INVOLVED IN
TRANSCRIPTIONAL REGULATION OF P21 IN RESPONSE TO DNA DAMAGE (BY
SIMILARITY).
CC -!- SUBUNIT: CTIP INTERACTS SPECIFICALLY WITH THE BRC1 DOMAINS (BY
SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Nuclear (Potential).
CC -!- TISSUE SPECIFICITY: IN THE EMBRYO, EXPRESSED IN OTIC VESICLES AT
DAY 9.5. AT DAY 10.5, THIS EXPRESSION DECREASES AND HIGH LEVELS
ARE FOUND IN THE NEUROECTODERM. AT DAYS 11-12.5, HIGH LEVELS IN
DIFFERENTIATING KERATINOCYTES AND WHISKER PAD PRIMORDIA. AT DAYS
14-17, EXPRESSION ALSO OBSERVED IN KIDNEY EPITHELIAL CELLS. IN
THE ADULT, HIGHEST LEVELS FOUND IN SPLEEN, THYMUS, LYMPH NODES,
EPITHELIAL ORGANS, AND ALVEOLAR AND DUCTAL EPITHELIAL CELLS OF
THE MAMMARY GLAND. VERY LOW LEVELS IN BRAIN, KIDNEY, AND SKIN. NO
EXPRESSION IN HEART, LIVER OR LUNG.


```
Qy 1047 GGSVPREFTGIIIPSTHELPMGLSGSSSAVQHQDDTQHNOQDEMKASHLQKTFELD 1106
Db 1440 GPTSGDDESGMRFS-PFKSPLAGRSASGCSRHLOKRNSPSEELLQAPG----- 1490
Qy 1107 LNSSEECITROSSTKONIT-DGCLPRORTADVDVPLNNSSLQNLIVLSSNNKEQTAV 1165
Db 1491 -----SEASSEPHNSTGSCPLR-RELEG-TPYLGSGLFSLFSDRDPESSEKPAHI 1540
Qy 1166 EYKTNATILREMGTADGKKPTSQWD-----SLRKDVENEGRQERNKNMDS-- 1215
Db 1541 GTTTPASTSVLKPOQVAFRAAAAGADKAVGVIVSKIKPELTSEERADRIDSMVYSG 1600
Qy 1216 IDYCAIRRASISE-----ISEAIKRGNNMLAVRIKDFLERIVKDHGGIDLEWRESPP 1270
Db 1601 TPKEVTVQKFAEYKRLTLDAITEETHVLIKDAEFCERTLK----- 1645
Qy 1271 DKADYLLSIRG-----IGLKSVECVRLTLHLNLAFFVDVTNVR 1309
Db 1646 -----YFLGIAGGWIVSYSWVRSIQERLLNVHEFEVTGDVVVTGR 1687

RESULT 13
CBP_HUMAN
ID CBP_HUMAN STANDARD: PRT; 2442 AA.
AC Q92793; Q16376; O00147;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE CREB-binding protein.
GN CREBBP OR CBP.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1]
SEQUENCE FROM N.A.
MEDLINE=97385172; PubMed=9238046;
RA Sobulo O.M., Borow J., Tomek R., Reshimi S., Harden A.,
RA Schlegelberger B., Housman D., Doggett N.A., Rowley J.D.,
RA Zeleznik-Le N.J.;
RT "MLL is fused to CBP, a histone acetyltransferase, in therapy-related
RT acute myeloid leukemia with a t(11;16)(q23;p13.3).";
RL Proc. Natl. Acad. Sci. U.S.A. 94:8732-8737(1997).
[2]
SEQUENCE FROM N.A.
MEDLINE=97321049; PubMed=9177780;
RA Giles R.H., Petrij F., Dauwerse H.G., den Hollander A.I.,
RA Lushnikova T., van Ommen G.J.B., Goodman R.H., Deaven L.L.,
RA Doggett N.A., Peters D.J.M., Breuning M.H.;
RT "Construction of a 1.2-Mb contig surrounding, and molecular analysis
RT of, the human CREB-binding protein (CBP/CREBBP) gene on chromosome
RT 16p13.3.";
RL Genomics 42:96-144(1997).
[3]
SEQUENCE OF 1-405 FROM N.A.
MEDLINE=96376968; PubMed=8782817;
RA Borow J., Stanton V.P., Andresen J.M., Becher R., Behm F.G.,
RA Chaganti R.S.K., Clivin C.I., Distche C., Dube I., Frischauf A.M.,
RA Housman D., Mittleman F., Volinia S., Watmore A.E., Housman D.E.;
RT "The translocation t(8;16)(p11;p13) of acute myeloid leukaemia fuses
RT a putative acetyltransferase to the CREB-binding protein.";
RL Nat. Genet. 14:33-41(1996).
CC -!- FUNCTION: MEDIATES CAMP-GENE REGULATION BY BINDING SPECIFICALLY TO
CC PHOSPHORYLATED CREB PROTEIN. ACTING AS A COACTIVATOR, CBP AUGMENTS
CC THE ACTIVITY OF PHOSPHORYLATED CREB TO ACTIVATE TRANSCRIPTION OF
CC CAMP-RESPONSIVE GENES.
CC -!- SUBUNIT: INTERACTS WITH SMAD1, SMAD2 AND SMAD3.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- DISEASE: INVOLVED IN ACUTE LEUKEMIAS BY CHROMOSOMAL TRANSLOCATIONS
CC T(8;16)(p11;p13) INVOLVING CBP AND MOZ, AND T(11;16)(Q23;P13.3)
CC INVOLVING CBP AND MLL.
CC -!- DISEASE: DEFECTS IN CREBBP ARE THE CAUSE OF RUBINSTEIN-TAYBI
```

```
CC SYNDROME (RTS), A DISORDER CHARACTERIZED BY CRANIOFACIAL
CC ABNORMALITIES, BROAD THUMBS, BROAD BIG TOES, MENTAL RETARDATION
CC AND A PROPENSITY FOR DEVELOPMENT OF MALIGNANCIES.
CC -!- SIMILARITY: CONTAINS 1 BROMODOMAIN.
CC -!- SIMILARITY: CONTAINS 1 ZN-TYPE ZINC FINGER.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement. Usage by and for commercial
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; UA7741; AAC51770.1; -
CC EMBL; U85962; AAC51331.1; -
CC EMBL; U89354; AAC51339.1; -
CC EMBL; U89355; AAC51340.1; -
CC MIM; 600140; -
CC InterPro: IPR001487; Bromodomain.
CC InterPro: IPR003101; KIX.
CC InterPro: IPR000197; TAZ_finger.
CC InterPro: IPR000433; Znf_ZZ.
CC Pfam; PF00439; bromodomain; 1.
CC Pfam; PF02135; KIX; 1.
CC Pfam; PF02172; zf-TAZ; 2.
CC PRINTS; PR00569; ZZ; 1.
CC PRINTS; PR00503; BROMODOMAIN.
CC SMART; SM00297; BROMO; 1.
CC PROSITE; PS00633; BROMODOMAIN_1; 1.
CC PROSITE; PS00014; BROMODOMAIN_2; 1.
CC PROSITE; PS01357; ZF_ZZ_1; 1.
CC PROSITE; PS01355; ZF_ZZ_2; 1.
KW Transcription regulation; Nuclear protein; Activator; Bromodomain;
KW Chromosomal translocation; Zinc-finger.
FT ZN_FING 1701 1744
FT DOMAIN 363 430 CYS/HIS-RICH.
FT DOMAIN 452 683 CREB-BINDING.
FT DOMAIN 1103 1175 BROMODOMAIN.
FT DOMAIN 1061 1064 POLY-GLU.
FT DOMAIN 1199 1487 CYS/HIS-RICH.
FT DOMAIN 1555 1562 POLY-GLU.
FT DOMAIN 1675 1849 CYS/HIS-RICH.
FT DOMAIN 1943 1948 POLY-PRO.
FT DOMAIN 1967 1970 POLY-GLN.
FT DOMAIN 2081 2085 POLY-GLN.
FT DOMAIN 2199 2216 POLY-GLN.
FT DOMAIN 2245 2248 POLY-GLN.
FT DOMAIN 2297 2300 POLY-GLN.
FT CONFLICT 1511 1513 FAE -> NSG (IN REF. 2).
FT CONFLICT 1724 1725 ED -> VV (IN REF. 2).
FT CONFLICT 1770 1770 V -> L (IN REF. 2).
FT CONFLICT 1789 1789 N -> F (IN REF. 2).
FT CONFLICT 1812 1812 T -> P (IN REF. 2).
SQ SEQUENCE 2442 AA; 265336 MW; 42D084619475F3D2 CRC64;

Query Match 2.1%; Score 192; DB 1; Length 2442;
Best Local Similarity 18.1%; Pred. No. 0.038;
Matches 345; Conservative 238; Mismatches 720; Indels 598; Gaps 89;

Qy 5 MDSSAVNATATEQDNGSRQDVLEFDLTKT-----POQPSKRRKRPMPKVVVEGKPKRK 59
Db 288 MGATGVNPQLASKQSMVNSLPTFTPTDIKNTSVTNVPMNSQMTSVGVPTQATATGPTAD 347
Qy 60 P--RKPAELPKVVVEGPKRKRKAATQEKVKSKTGSAKKNLKSATK----KPAENVG 113
Db 348 PEKRLKQQQLVLLHLLHAKCORREQANGEVACSLPHCRRTMKVNLNHTHCQACKAQA 407
Qy 114 DMSNKSPEVT-LKSCRK---ALNFDLENPGDARQGD----SESEIVQNSGANSFSFIRD 165
Db 408 HCASSRQIISHWNKTRHDCPVCPLPKNASDKRKNQQTILGSPASGIQNTIGS-----V 460
```

```
QY 166 AIGGTNGSFLSVSVIDKTN-----GLGAMNQPLEV----- 196
D 166 AIGGTNGSFLSVSVIDKTN-----GLGAMNQPLEV----- 196
D 461 GTGQONATSLSNPNP IDPSSMORAYAAALGLPYMNPOTQLQPVGQOPAPQTHQOQRT 520
QY 197 --SMGNQPKLSTGAKLARDQDPLLTNRNOQQCPVATONTQFPMENQOQWLQKNQNLIG 254
D 197 --SMGNQPKLSTGAKLARDQDPLLTNRNOQQCPVATONTQFPMENQOQWLQKNQNLIG 254
D 521 LNPLGNPNWNPAGG-ITDQOQPNLISEALPTSLGATN-----PLMND-----G 565
QY 255 PFGNQOQPMTRIRNOQPCLAGMNOQPMV-----LIGTPRPA----- 290
D 255 PFGNQOQPMTRIRNOQPCLAGMNOQPMV-----LIGTPRPA----- 290
D 566 SNSNGNIGTLSTIPRAAPSSGTGRGWEHVHTQDLRSHLVHKLVAIPTPDPAALKDR 625
QY 291 ---LVSGNQQLGG---PGNKRRIPLNHOTCLPAGNQLYGSPTDQHLQVMTGGQOQGLL 344
D 291 ---LVSGNQQLGG---PGNKRRIPLNHOTCLPAGNQLYGSPTDQHLQVMTGGQOQGLL 344
D 626 MENLVAYAKKVEGDMYESSANSRDEYH-----LLAEKIYKIQKELEEKRRS---RLHQOG 677
QY 345 IKNOQPGSLIRGOQPCVPLIDQ-OPATPKGFTHLNQWVATSNSSPGLRPHQSQVPTTYL 403
D 345 IKNOQPGSLIRGOQPCVPLIDQ-OPATPKGFTHLNQWVATSNSSPGLRPHQSQVPTTYL 403
D 678 ILGNOPALPAPGAQD--PVIPAQVPRP-----PNGPLSLPVNRM 715
QY 404 HVESVSRILNG--TTGTCQSRAPAYDSLQODIHOGNKYILSHEISNGNGCKKALPONS 462
D 404 HVESVSRILNG--TTGTCQSRAPAYDSLQODIHOGNKYILSHEISNGNGCKKALPONS 462
D 716 QVSGMNSFNPSLGNVQLPOAPM-----GPRAAAPMNSHV 751
QY 463 PTPIMAKLEEARGSKQYHRAGQTEKHDLN-LAQQIAQSQDVERHNSSTCVELDA--- 518
D 463 PTPIMAKLEEARGSKQYHRAGQTEKHDLN-LAQQIAQSQDVERHNSSTCVELDA--- 518
D 752 QMNSMGVPGMAISPMPQPPNMGAHTNNMAQAPASQFLPQNPSPSSGAMSVGMG 811
QY 519 -----AKTKIKQVQENLH----- 533
D 519 -----AKTKIKQVQENLH----- 533
D 812 QPPAQGTGVSQGVGAALPNLMLGQASQLPCPVVTSPLHTPPPPASTAAGMPSLOH 871
QY 534 ---GMPPEVIEIEDPDGARKKNTASISKG-----ASKGNSSP-VKKTAEKEKCIYVK 584
D 534 ---GMPPEVIEIEDPDGARKKNTASISKG-----ASKGNSSP-VKKTAEKEKCIYVK 584
D 872 TTPQMTTPQPAATPQSTPVSSSQOTPTPTPGSVPSATQSTPTVQAAQAQVTPQO 931
QY 585 TPAAKGRAGRKS---VPPPAH-----ASEIOLMQPTP-----PK 616
D 585 TPAAKGRAGRKS---VPPPAH-----ASEIOLMQPTP-----PK 616
D 932 TPVQPPSVATPQSSQOQPTPVHAQPGTGPLSOAAASIDNRVPTPSSVASAETNSQQPGD 991
QY 617 TPLRSKPKGGRKSIQDSKARG-PSGELLCDOSIAELIYRMQNLXLGDKEREQEQNAM 675
D 617 TPLRSKPKGGRKSIQDSKARG-PSGELLCDOSIAELIYRMQNLXLGDKEREQEQNAM 675
D 992 VPVLEMTQETQEDTPDGGSGEPSEMEED-----LOGASQVKEETDI 1038
QY 676 VLYKDGALVPYSEKKRPRKVDIDDTTRIWLLMLGKDEGEDEKDKKKKEKWDEE 735
D 676 VLYKDGALVPYSEKKRPRKVDIDDTTRIWLLMLGKDEGEDEKDKKKKEKWDEE 735
D 1039 AEQKSE---PMEVDEKKPEYKVEKEEESNN---GTASQSTSPSQPRKKIKF-PEEL 1090
QY 736 RRVFGRADSFIAHMLVQGRDRFPSPWKGVSVDVYGVFLTQNVSDHLSGSAFMSLAARF 795
D 736 RRVFGRADSFIAHMLVQGRDRFPSPWKGVSVDVYGVFLTQNVSDHLSGSAFMSLAARF 795
D 1091 QOALMPTLEA-----LYRQDPESLIPFRQPDVQLLGI---PDVFDIVKPNMDLSTIKR- 1140
QY 796 PPKLSSSREDE--RNVRSVVVEDEGCILN--LNEIPSWQEKVQHPSPDMEVS-----GV 845
D 796 PPKLSSSREDE--RNVRSVVVEDEGCILN--LNEIPSWQEKVQHPSPDMEVS-----GV 845
D 1141 --KLDGTQYQEPWQVDDVWLFMNNAWLNKRTSKVYKFCSKLAEVFFQEDIPVWQSLUGY 1198
QY 846 DSGSK-----BQL-----RDCNSGIE-RFNLEKSIQNLLEVLN----- 880
D 846 DSGSK-----BQL-----RDCNSGIE-RFNLEKSIQNLLEVLN----- 880
D 1199 CCGRYEFPOTLCYQKQLCTIPRDAAYSYQNYRHFCEKCFEIQEIGNVTLGLDDPSQ 1258
QY 881 -----SQDSFPDPAIFQSGRRVGS----- 1258
D 881 -----SQDSFPDPAIFQSGRRVGS----- 1258
D 1259 QTTISKDOFEKKKNDLDPPEVDCKEGGRMHQICVLHYDIIWPSGFCVDCNCLKTKGR- 1317
QY 909 PTTRECKTVSGTSQSVGTGSPNLSDEI--CLOGNERPHLYEGSGDV-----QKQETTN 960
D 909 PTTRECKTVSGTSQSVGTGSPNLSDEI--CLOGNERPHLYEGSGDV-----QKQETTN 960
D 1318 --PRKNKFSAKRLQTTNLGN-HLEDNRVKNFLRRQNHEP-----AGEVVRVVASDKTVE 1370
QY 961 VAQKPKDLEKTMNWKDSVCFQPRNDTWNQTTTPSSSYEQCATRQPHVLVDIEDFGMOGBGL 1020
D 961 VAQKPKDLEKTMNWKDSVCFQPRNDTWNQTTTPSSSYEQCATRQPHVLVDIEDFGMOGBGL 1020
D 1371 V---KPGMKS--RFVDS---GEMSESFYPTKALFAFEEI-----DGVDCVCFGMHVOEY 1417
```

```
QY 1021 GYSWMSISPRVDRVKNKNVPR-RFFRQGGSVPR-----EF-----TGQ 1057
D 1021 GYSWMSISPRVDRVKNKNVPR-RFFRQGGSVPR-----EF-----TGQ 1057
D 1418 GSD--CPPNTRRVYISYLSIHFFR-----PRLRTAVYHEILIGYLEYVKLGIVTGH 1470
QY 1058 IIPSTPHELPCMGSLGSSSAVQEHQDDTHQHQDQEMKNASHLOKTFDLNLS--SEECUT 1115
D 1058 IIPSTPHELPCMGSLGSSSAVQEHQDDTHQHQDQEMKNASHLOKTFDLNLS--SEECUT 1115
D 1471 IWACPPSE-----GDDYIFHCAPPD-----OKIPKPKRLQEWYKMLDKAFARIH 1517
QY 1116 QSSSTKQNTDGLCLPRDRTAEDVVDPDLSNNSLQNLIVESNSSNKEQTAQVAYKETNATIL 1175
D 1116 QSSSTKQNTDGLCLPRDRTAEDVVDPDLSNNSLQNLIVESNSSNKEQTAQVAYKETNATIL 1175
D 1518 DYKIDFKQATE-----DLRTSAKELPYEGDFWPNVLESIKLEQEBEERKKESTAAS 1572
QY 1176 REMKGTIADGKKPTSQWDSLRKQVEGROERKNKNNMNSIDYBAIRASISSEISAEIK 1235
D 1176 REMKGTIADGKKPTSQWDSLRKQVEGROERKNKNNMNSIDYBAIRASISSEISAEIK 1235
D 1573 ETTEGSGQDSK-----NAKKNNKNTKNK-----SSISRANKK 1606
QY 1236 R---GNNMLAVRIKOFLEP-----IVKDHGGIDLEWLRSPDKAKDYLLSIRGLGL 1285
D 1236 R---GNNMLAVRIKOFLEP-----IVKDHGGIDLEWLRSPDKAKDYLLSIRGLGL 1285
D 1607 KPSMPNVSNDLSQKLYATMEKHKEVFFVIHLHAGPVINTL---PPIVDPDPLLS----- 1657
QY 1286 KSEVCVRLTLHLNLAFFVDVTNVGR-----IAVRMGWVPLQLPESLQLHLLELYPVLES 1340
D 1286 KSEVCVRLTLHLNLAFFVDVTNVGR-----IAVRMGWVPLQLPESLQLHLLELYPVLES 1340
D 1658 -----CDLMDGRDAFLTLDARKHW-----EFSSL 1681
QY 1341 QKFLWPRCLKDQRTLYELHYQLITFGK---VFCTKSRPNCNACPMRGECRHFASAYASA 1397
D 1341 QKFLWPRCLKDQRTLYELHYQLITFGK---VFCTKSRPNCNACPMRGECRHFASAYASA 1397
D 1682 RRSKWSLTLC---MLVELHTQ---GQDREYV-----TCNECKHHVETRWHTCTVCEDY 1726
QY 1398 RLALPAPERSLTSATIPVPESPPVPAIPMLPLPL-EKSLGAPSNRENCEPIIEE 1456
D 1398 RLALPAPERSLTSATIPVPESPPVPAIPMLPLPL-EKSLGAPSNRENCEPIIEE 1456
D 1727 DLICNICYNTKS-----HAHKMWKGLDDEGSSQGEPOSK----- 1762
QY 1457 PASPQCECTEIT-----ESDIEDAYNEDDDEIPTIKLNIQFGMTLREHMERNNMLOEG 1511
D 1457 PASPQCECTEIT-----ESDIEDAYNEDDDEIPTIKLNIQFGMTLREHMERNNMLOEG 1511
D 1763 --SP-QESRRYSIQRCIQSLVHACQCRNANCSLPSC-----QKKRVVQHTPKG 1807
QY 1512 -----DMSKALVAL-----HPTTTSIPTKLNKIS-RLR---TEHOVYVELDPSHR 1552
D 1512 -----DMSKALVAL-----HPTTTSIPTKLNKIS-RLR---TEHOVYVELDPSHR 1552
D 1808 CKRKTNGGCPVCVKQIILACCYHAKHCKQENKCPVPFCLNKKIKLRQQOIQHRLQQAQLMRR 1867
QY 1553 LLDGMDKR---EPDDPSPYLLAIWTPGE---TANSAQOPPEQ 1587
D 1553 LLDGMDKR---EPDDPSPYLLAIWTPGE---TANSAQOPPEQ 1587
D 1868 RMAWTRNTRVQOQSLPSTAPPSTPQOQSTPOTPQPPAQ 1908
RESULT 14
TPR_HUMAN
ID TPR_HUMAN STANDARD; PRT: 2349 AA.
AC P12270;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Nucleoprotein TPR.
GN TPR.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93064711; PubMed=1437155;
RA Mitchell P.J., Cooper C.S.;
RT "The human tpr gene encodes a protein of 2094 amino acids that has
RT extensive coiled-coil regions and an acidic C-terminal domain.";
RL Oncogene 7:2329-2333(1992).
RN [2]
RP REVISIONS, AND CHARACTERIZATION.
RX MEDLINE=95096166; PubMed=7798308;
RA Byrd D.A., Sweet D.J., Pante N., Konstantinov K.N., Guan T.,
RA Saphire A.C.S., Mitchell P.J., Cooper C.S., Aebi U., Gerace L.;
RT "Tpr, a large coiled coil protein whose amino terminus is involved in
RT activation of oncogenic kinases, is localized to the cytoplasmic
RT surface of the nuclear pore complex.";
```


Db 935 YDEAREEVRLRQELTQL-HEDLKRARESDKNDYSYKRELERQRAKVIEVEKELEERYFD 993
QY 1256 D-----HGGIDLEWLRESPPDKADYLLSIR--GLGLKSVECVRLLLPLHNLAFPVDT 1305
Db 994 DSRLAECKRHGD---EMLR-----KIADLEKLRDGGNGGNGCTSSCEFER----- 1038
QY 1306 NVGRIAVRMGWPLQPLPESLQHLLELYPVLESIQKFL-WPRL---C--KLDQ----- 1353
Db 1039 --KRIAV-----LEVEVRKSMETIKSLEKFMEDRLQDKCADKLDREKERRM 1083
QY 1354 RTLYELHYQLITFGKVFTKSRPNCNACPMRGECRHFASAYASARLALPAPEE----- 1406
Db 1084 KAERDLEREI-----ARNCNGNCPERELESERSNVKREYOLDAEKVKFYKR 1133
QY 1407 -----RSLTSATIPVPPEFPVPAIPMIELPLPLEKSLASGAPSNRENCEPIIEE 1456
Db 1134 ELERDYLSSRYLTSSDP-DEKPLPNYTFPRIV-----EPLTTEDEPKPVE 1181
QY 1457 PASPGQECTEITESDIEDAYNEDDDEIPTIKLNIEQFGMTLREHMERNMELQEGDMSKA 1516
Db 1182 WPPSSDVTEPISSGV-----TPSVDAEPE-----HPQLSEYQTSVS 1218
QY 1517 LVALHPTTTSIPTPKLKNISRLRTEHQVYELPD SHR-----LLDGMKREPDPS 1567
Db 1219 QVAVTTP-----PKPETQISEYQDSELYSASNNTESKNVFSLAYLDDLDKL--DDIDE 1272
QY 1568 YLLAIWTPGET 1578
Db 1273 YLLNNIMPEKT 1283

Search completed: September 27, 2002, 13:14:22
Job time: 1110 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 27, 2002, 12:54:47 ; Search time 122.1 Seconds
(without alignments)
2449.699 Million cell updates/sec

Title: US-09-840-743-2

Perfect score: 9089

Sequence: 1 MOSIMSSAVNATEQND.....PRPLMARLHPASKLNKNT 1729

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_19:**

- 1: sp_archaea:**
- 2: sp_bacteria:**
- 3: sp_fungi:**
- 4: sp_human:**
- 5: sp_invertebrate:**
- 6: sp_mammal:**
- 7: sp_mhc:**
- 8: sp_organelle:**
- 9: sp_phase:**
- 10: sp_plant:**
- 11: sp_rodent:**
- 12: sp_virus:**
- 13: sp_vertibrate:**
- 14: sp_unclassified:**
- 15: sp_rvirus:**
- 16: sp_bacteriap:**
- 17: sp_archaeap:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	3949	43.4	1017	Q9L269	Q9L269 arabidopsis
2	2863	31.5	555	Q9L268	Q9L268 arabidopsis
3	1776.5	19.5	1309	Q9SR66	Q9SR66 arabidopsis
4	1559.5	17.2	1207	Q9SJG6	Q9SJG6 arabidopsis
5	1260	13.9	234	Q9L267	Q9L267 arabidopsis
6	999	11.0	274	Q9FTQ2	Q9FTQ2 oryza sativ
7	709	7.8	917	Q9498	Q9498 arabidopsis
8	458	5.0	119	Q94LX6	Q94LX6 euphorbia t
9	281	3.1	2151	Q9NG79	Q9NG79 trichomonas
10	278	3.1	5327	Q96891	Q96891 drosophila
11	253	2.8	17352	Q95YM2	Q95YM2 procambiarus
12	245	2.7	2254	Q9HCY0	Q9HCY0 homo sapien
13	241	2.7	2187	Q9H197	Q9H197 homo sapien
14	240.5	2.6	2938	Q61769	Q61769 mus musculu
15	239	2.6	2701	Q9Y520	Q9Y520 homo sapien
16	239	2.6	3574	Q9AUB4	Q9AUB4 arabidopsis

ALIGNMENTS

RESULT 1

Q9L269 ID Q9L269 PRELIMINARY; PRT; 1017 AA.
AC Q9L269;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-OCT-2000 (Tremblrel. 15, Last annotation update)
DE HYPOTHETICAL I12.1 KDA PROTEIN.
GN T32M21_160.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_taxid=3702;
RN SEQUENCE FROM N.A.
RP Bevan M., Terry N., Ardiles W., Buysshaert C., Dasseville R.,
RA De Clerck R., De Keyser A., Neyt P., Rouze P., Van Den Daele H.,
RA Villarroel R., Gielen J., Van Montagu M., Bancroft I., Mewes H.W.,
RA Rudd S., Lemcke K., Mayer K.F.X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL162875; CAB8562.1; -
KW Hypothetical protein.
SQ SEQUENCE 1017 AA; 4F7C313A891EC83F CRC64;

Query Match 43.4%; Score 3949; DB 10; Length 1017;
Best Local Similarity 99.9%; Pred. No. 1.8e-246;
Matches 755; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MOSIMSSAVNATEQNDGSRQDVLEFDLNKTPQKPKRKFMPKVVVEGPKRKP 60

Db 235 MOSIMSSAVNATEQNDGSRQDVLEFDLNKTPQKPKRKFMPKVVVEGPKRKP 294

QY 61 RKPAELPKVVVEGPKRKPRAATQEKVKSETGSAKKNLKESATKKPANVGDMNSKP 120

Db 295 RKPAAELPKVVEGPKRKPRAKQAEYKSKETSAKKNLKESATKKPANVGDMNSKSP 354
Qy 121 EVTLKSCRKALNFDLEPGDARQDSSEIIVONSSGANSFSEIRDAIGGTNGSFLDSVSQ 180
Db 355 EVTLKSCRKALNFDLEPGDARQDSSEIIVONSSGANSFSEIRDAIGGTNGSFLDSVSQ 414
Qy 181 IDKTNGLGAMNOPLEVSNGNPDKLSTGAKLARDQDQDLLTRNQCCQFPVATQNTQPFME 240
Db 415 IDKTNGLGAMNOPLEVSNGNPDKLSTGAKLARDQDQDLLTRNQCCQFPVATQNTQPFME 474
Qy 241 NOQAWLQMKNLIGFPNGNQPRWTIRNOQPCLAGNQOQMYLIGTPRPALVSGNQQLGG 300
Db 475 NOQAWLQMKNLIGFPNGNQPRWTIRNOQPCLAGNQOQMYLIGTPRPALVSGNQQLGG 534
Qy 301 POGNKRPIFLAHQTHCLPAGNOLYGSPTDMHOLVMTSGQQHGLLIKNOOPGSLIRGOQPC 360
Db 535 POGNKRPIFLAHQTHCLPAGNOLYGSPTDMHOLVMTSGQQHGLLIKNOOPGSLIRGOQPC 594
Qy 361 VPLIDQOQPATPKGFTHLNQMVATSMSSPGLRPHSQSQVPTTYLHVESVSRILNGTTGTCQ 420
Db 595 VPLIDQOQPATPKGFTHLNQMVATSMSSPGLRPHSQSQVPTTYLHVESVSRILNGTTGTCQ 654
Qy 421 RSRAPAYDSLOODIHQGNKYILSHSISNGNCKKALPONSSSLPTPIAKLEEARGSRKQY 480
Db 655 RSRAPAYDSLOODIHQGNKYILSHSISNGNCKKALPONSSSLPTPIAKLEEARGSRKQY 714
Qy 481 HRAMGOTEKHLNLAQQAQIAQSQDVERHNSSTCEVYLDAAKTKTKQVVOENLHGMPPPEVI 540
Db 715 HRAMGOTEKHLNLAQQAQIAQSQDVERHNSSTCEVYLDAAKTKTKQVVOENLHGMPPPEVI 774
Qy 541 ETEDDPTDGARKGNKNTASISGASKGNSPVKKTAKEKCIIVPKTPAKKGRAGRKKSVP 600
Db 775 ETEDDPTDGARKGNKNTASISGASKGNSPVKKTAKEKCIIVPKTPAKKGRAGRKKSVP 834
Qy 601 PAHASEIQLWQPTPKTPLSRKPKGKRKSTQDSGKARGPSGELLCCODSTAEEIYRMON 660
Db 835 PAHASEIQLWQPTPKTPLSRKPKGKRKSTQDSGKARGPSGELLCCODSTAEEIYRMON 894
Qy 661 LYLGGKEREQEQANVLYKGDGALVPYESSKKRPKPKYDIDDETTIRIWNLLMGKDEKEG 720
Db 895 LYLGGKEREQEQANVLYKGDGALVPYESSKKRPKPKYDIDDETTIRIWNLLMGKDEKEG 954
Qy 721 DEEDKKKEKWEERRVFRGRADSFIAHMLVQGD 756
Db 955 DEEDKKKEKWEERRVFRGRADSFIAHMLVQGE 990

RESULT 2
Q9LZ68 PRELIMINARY; PRT; 555 AA.

AC Q9LZ68;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DE 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE HYPOTHETICAL 62.1 KDA PROTEIN.
GN T32M21.170.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Bevan M., Terry N., Ardiles W., Buysshaert C., Dasseville R.,
RA De Clerck R., De Keyser A., Neyt P., Rouze P., Van Den Daele H.,
RA Villaroel R., Gielen J., Van Montagu M., Bancroft I., Mewes H.W.,
RA Rudd S., Lemcke K., Mayer K.F.X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL; AL162875; CAB85563.1; -.
DR InterPro; IPR003265; Endo_3c.
DR SMART; SM00478; ENDO3c; 1.
KW Hypothetical protein.
SQ SEQUENCE 555 AA; 62088 MW; A1B44BDDEL17FDC1E CRC64;

Query Match 31.5%; Score 2863; DB 10; Length 555;
Best Local Similarity 100.0%; Pred. No. 1.1e-176;
Matches 549; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 785 SSAPMSLAARPPPKLSSREDERNVRSVVVEDPEGCIILNLEIPSWOEKVKVQHPDSMEVSG 844
Db 7 SSAPMSLAARPPPKLSSREDERNVRSVVVEDPEGCIILNLEIPSWOEKVKVQHPDSMEVSG 66
Qy 845 VDSGSKQLRDCNSNGIERFNFLFKSIQNLEEEVSSQDSFDPALFQSCGRVSGSCSKS 904
Db 67 VDSGSKQLRDCNSNGIERFNFLFKSIQNLEEEVSSQDSFDPALFQSCGRVSGSCSKS 126
Qy 905 DAEPPTTRCETKTVSGTSQSQSVOTGSPNLSDICIQGNERPHLYEGSGDVQKQETTINVAOK 964
Db 127 DAEPPTTRCETKTVSGTSQSQSVOTGSPNLSDICIQGNERPHLYEGSGDVQKQETTINVAOK 186
Qy 965 KPDLEKTMNWKDSVCFGQPRNDTNWQTPSSSYEQCATRQPHVLDIEDFGMQGEGLGYSW 1024
Db 187 KPDLEKTMNWKDSVCFGQPRNDTNWQTPSSSYEQCATRQPHVLDIEDFGMQGEGLGYSW 246
Qy 1025 MSISPRVDRVKNKNVPRFRFQGGSVPREFTGQIIPSTPHELPGMGLSGSSAYQEHQDD 1084
Db 247 MSISPRVDRVKNKNVPRFRFQGGSVPREFTGQIIPSTPHELPGMGLSGSSAYQEHQDD 306
Qy 1085 TOHNOQDEMKNKASHLQKTFDLNLSSECLTRQSTKQNTIDGCLPRDRTAEDVVDPLSN 1144
Db 307 TOHNOQDEMKNKASHLQKTFDLNLSSECLTRQSTKQNTIDGCLPRDRTAEDVVDPLSN 366
Qy 1145 NSSLQNTILVESNSSNKEQTAVEYKETNATILREMGKTLADGKPTSQWDSLRLKDVQEGNEG 1204
Db 367 NSSLQNTILVESNSSNKEQTAVEYKETNATILREMGKTLADGKPTSQWDSLRLKDVQEGNEG 426
Qy 1205 ROERNKNMDSIDYEAIRRASISEISEAIKERGMNMLAVRIKDFLERIVKHGIDLEW 1264
Db 427 ROERNKNMDSIDYEAIRRASISEISEAIKERGMNMLAVRIKDFLERIVKHGIDLEW 486
Qy 1265 LRESPDPKADYLLSIRGLGLKSVCEVRLTLHLNLAFPVDTNVGRIAVRMGNVLPLOLPE 1324
Db 487 LRESPDPKADYLLSIRGLGLKSVCEVRLTLHLNLAFPVDTNVGRIAVRMGNVLPLOLPE 546
Qy 1325 SIQLHLEL 1333
Db 547 SIQLHLEL 555

RESULT 3
Q9SR66 PRELIMINARY; PRT; 1309 AA.
AC Q9SR66;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE T22K18.18 PROTEIN.
GN T22K18.18.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Lin X., Kaul S., Town C.D., Benito M., Creasy T.H., Haas B.,
RA Ronning C.M., Koo H., Fujii C.Y., Utterback T.R., Barnstead M.E.,
RA Bowman C.L., White O., Nierman W.C., Fraser C.M.;
RT "Arabidopsis thaliana chromosome III BAC T22K18 genomic sequence.";

Qy	73	GPKPRKRAAQTQEVK--SKETGSAKKKNLKESATTKKPANVGMDSNKSPEVTLKSGRKAL	131
Db	121	GOSEKTPKRYKRVKKEVSKD-----QDATPVESSAAVETSRPK---RLCRVVL	167
Qy	132	NFDLEN----PGDARQ--GDSESEIVQNSGANSFSELRDAIGTNGSELDVSQIDKTN	185
Db	168	DFAENGENTGDIIEAGEMESAEOEKOLDG--QELUKCL-----LSAFSTPKRK	219
Qy	186	GLGAMQPLEVSMGNQDPK-----LSTGAKLARDOQPDLLTRN-----QOCQFP	229
Db	220	SQGRK-----GVQPKNGSNLEEVDISMAQAARQRPCTCCDMNLGQIDYEQCDY-	271
Qy	230	VATONTQFPMENQOAWLQMKNOLOGIPFPGNOOPRWTIRNOQPCIAMGNQOPMYLIGTPRP	289
Db	272	-----QKMHWLSPNL-----QOGGM--RYDAIC-----	293
Qy	290	ALYSGNOQLGGPQGNKRPFILFHNQTCLPAGNOLYGSPTDMHQLYMSITGGQOGLLIKNOQ	349
Db	294	-----SKVFSGQGHNYVSAF--HATCYSTSOLSANRV-----LTVEERR	331
Qy	350	PGSLRIGQOPCVPLIDQOQATP-----KGFTHLQNMVATSMSSPGLRPHSQS	398
Db	332	EGIFQGRQSELSNVLSDKIDTPIKKKTHCHARFNLSSMKNLV-----EV	376
Qy	399	PTYTLHVESVSRILNTTGTQORSAPAYDSLOQDIHOGNKYILSHEISNGNGCKKALPQ	458
Db	377	PE---HL-----TSGYCSKPQ-----ONKKLVDTRVT-----	401
Qy	459	NSSLPTPIAKLEEARGSKROYHRAMQGTQEKHDLNLAQOIAQSODVERHNSCTCVELDA	518
Db	402	-----VSKKPKTKSEKSO-----	414
Qy	519	AKTKTIQVQENLHGMPEVIEIEDDPTDGARKGNKTASISKASKGNSPVKKTAEKE	578
Db	415	--TK-QKNLLPNLCRFPFSFTGLSPD-----	437
Qy	579	KCIVPTPAKKGRAGRKKSVPYPAHASEIQUWQPTPKTPLSRKPKGKGRKSIQDSGKA	638
Db	438	-----ELW-----KRNST-----	446
Qy	639	RGPSGELLQDSTAEILYRMQNLYLGDKEREOENAWLYK-----GDGALVPYE	688
Db	447	-----ETISEL-----LRLLDINREHSETALVPYTNNSQIVLFGGAGAIYFVT	490
Qy	689	S-KRRKPRPKVDIDDETTIRWMLMKGDKEBGDEEKKKKEKWEERERRVFRGRADFSI	747
Db	491	PVKKPRPKVDLDDETDVRWKLILL-ENINSGVDGSEQAKKWEERNVFRGRADFSI	549
Qy	748	ARMHLVQDRRRSPKSGVSDSVIGVFLTQNVSDHLSSSAFWSLAARPP-PKLSSSREDE	806
Db	550	ARMHLVQDRRRTPWKSGVSDSVGVFLTQNVSDHLSSSAFWSLASOPFPVPFVPSNFD-	608
Qy	807	RNYSRVVVEDPECCIILNLEIFSWQ-----EKVQHFSDMEVSGVDSGSKQLRDCSNS	859
Db	609	-----AGTSMSPSIQITYLDSEETMSSPPDHNHSSVT-----	640
Qy	860	GIERFNPLEKSIONEEVLSSQDSFDPAFIOFCGRGVGSCSCSDAEFTTRCET--KT	917
Db	641	-----LKNQTPDEEKDVPVSNET-----SRSSSEIAISAHESVDKT	676
Qy	918	VSGTS--QSVQOTGSPNLSDBEICLQNERPHLYEGSGDVOK-OETTNAQKPKDLEKTMNW	974
Db	677	TDKEYVSDSRKGS-----SVEVDKTDKCRVLNLPSPSEDSALTC	716
Qy	975	KDSVCFOGPNNDTNWQTTSPSSYEQCATRQPHVLIDIEFGMQGELGYSNWSISPRVDRV	1034
Db	717	QHSMSVDAPONT---ERAGSSS-----EIDLEGE-YRFSFMKL-----	750
Qy	1035	KNKNVPRFRFGSGVSPREFTGQIIPSTPHELPGMLGSGSSSAVOEHODDTHQNOQDEM	1094
Db	751	-----LOGVOVSLSDSNQVSPN-----MSPGDCSSSEIKGFQ-----SMKEPT	787
Qy	1095	KASHLQKTFDLNLSSEE-CLTRQSGTKQNTDGLCLPRDRTAEDVVDPLSNNSLQNLIV	1153

	I : I	: :	: :	--DG-----DVL-----	807
Dn	788 KSS	-----VDSBPGCCSQO	:		
Qy	1154 ESNSNNKEQTAVYKKTNATILREMKGTIADCKPKTSOWDSLKDVEGNEGQRKNKM	1213			
Dn	808 ---SCQKP--LKGKKVKLE-----EKAFWDCLRLREAQAARAGIREKTRSTM	852			
Qy	1214 DSDYEAIRASISEIAIKERGMNMFLVRP-----KDFLERIVKHGGIDLE	1263			
Dn	853 DTVDWKAIARAADVKEVAETIKSGMNHKLAEIRIQYLTMNKIMOGFDRULVNDHGSIDL	912			
Qy	1264 WLRESPDKAKDYLLSTIRGLGLKSVECVRLTLHNLAFFVDITNVGRIAVRMGWVPLOPLP	1323			
Dn	913 WLRDVPPDKAK-----	923			
Qy	1324 ESLQLHLLEYLVLESIOKFLWPRCLKLDORTLYELHYOLITFGKVCTKSRPNCACPM	1383			
Dn	924 -----YELHYOMITFGKVCTKSPKNCACPM	950			
Qy	1384 RGEGRHFASFASARIALPAPEERSUTSATIPVPESPPPVAIMPDELPLEKSLASGA	1443			
Dn	951 KGEGRHFASFASARIALPSTEKMGTPDPKNLP-----LHLPFPQQEGSEV	999			
Qy	1444 PSNRN-----CEPIIEPASPGECTEITESIEDIAYNEDPDDEIPTIKLNIEQFGMT	1497			
Dn	1000 VQHSEPARKVTCCEPTIEEPASPETAESVIADIEEAFF-EDPEIPTIRLNDAFTSN	1058			
Qy	1498 LREHWERNMELOGDMSKALVALHPHTTISPTPKLKNSRRLTEHOVVYELPDSHRLDDGM	1557			
Dn	1059 LKKIMEINKELQDNWSALVALTAETSLLMPKLNLSOLETHRVVELDHEPLAQL	1118			
Qy	1558 DKREPDPDSPYLAIWTGPANSAPPQKCGGKASKMGCFDETCSNLSUREANSTGV	1617			
Dn	1119 EKREPDPSPYLAIWTGPETADSIQPSYSTCIGFANGMLCDEETFCSCNSIKETRISOIV	1178			
Qy	1618 RTLLI 1623				
Dn	1179 RGTIIV 1184				
RESULT	5				
Q9LZ67	PRELIMINARY;	PRT; 234 AA.			
ID	Q9LZ67				
AC	Q9LZ67;				
DT	01-OCT-2000 (TrEMBLrel. 15, Created)				
DT	01-OCT-2000 (TrEMBLrel. 15, Last sequence update)				
DT	01-OCT-2000 (TrEMBLrel. 15, Last annotation update)				
DE	HYPOTHETICAL 26.6 KDA PROTEIN.				
GN	T32M21.180.				
OS	Arabidopsis thaliana (Mouse-ear cress).				
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;				
OC	Spermatophyta; Magnoliopsida; eudicotyledons; core eudicots; Rosidae;				
OC	eurosid II; Brassicales; Brassicaceae; Arabidopsis.				
OX	NCBI_TaxId=3702;				
[1]					
RN	SEQUENCE FROM N.A.				
RA	Beyan M., Terryn N., Ardiles V., Buysshaert C., Dasserville R.,				
RA	De Clerck R., De Keyser A., Neyt P., Rouze P., Van Den Baele H.,				
RA	Villarroel J., Gielens J., Van Montagu M., Bancroft I., Mewes H.W.,				
RA	Rudd S., Lemcke K., Mayer K.F.X.;				
RL	Submitted (MAR-2000), to the EMBL/GenBank/DDBJ databases.				
[2]					
RN	SEQUENCE FROM N.A.				
RA	EU Arabidopsis sequencing project:				
RL	Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.				
DR	EMBL; AL162875; CAB85564.1; -				
KW	Hypothetical protein.				
SQ	SEQUENCE 234 AA; 26649 MW; B395392E06EB297 CRC64;				

Query Match 13.9%; Score 1260; DB 10; Length 234;
Best Local Similarity 100.0%; Pred. No. 1.2e-73;
Matches 234: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1496 MTLREHMERNMELQEGDMSKALVALHPTTTSIPTPKLNISRLRTERHQVYELPDSHRLD 1555
Db 1 MTLREHMERNMELQEGDMSKALVALHPTTTSIPTPKLNISRLRTERHQVYELPDSHRLD 60
QY 1556 GMDKREPDPSPYLLAIWTGCTANSQAQPPQKCGKASGKMCDFETCSECNLSREANSQ 1615
Db 61 GMDKREPDPSPYLLAIWTGCTANSQAQPPQKCGKASGKMCDFETCSECNLSREANSQ 120
QY 1616 TVRGTLIPCRMTARGSPFLNGTYFQVNELFADHESLKPIDVPRDWIWLPRRTVYFGT 1675
Db 121 TVRGTLIPCRMTARGSPFLNGTYFQVNELFADHESLKPIDVPRDWIWLPRRTVYFGT 180
QY 1676 SVTSIFRGLSTEQIQFCFWKGVCVRGFEQKTRAPRLMARLHPASKLNKNT 1729
Db 181 SVTSIFRGLSTEQIQFCFWKGVCVRGFEQKTRAPRLMARLHPASKLNKNT 234

RESULT 6
Q9FTQ2 PRELIMINARY; PRT; 274 AA.
AC Q9FTQ2
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE P0665D10.14 PROTEIN.
GN P0665D10.14
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzeae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. NIPPONBARE;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
RT clone:P0665D10."
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP002861; BAB16489.1; -.
SQ SEQUENCE 274 AA; 30951 MW; D7C826BD3982592D CRC64;

Query Match 11.0%; Score 999; DB 10; Length 274;
Best Local Similarity 68.6%; Pred. No. 1.1e-56;
Matches 179; Conservative 38; Mismatches 42; Indels 2; Gaps 2;

QY 1469 ESDIEDAYNEDPDPEITIKLNIEQFGWTLREHME-RNMELOEGDMSKALVALHPTTTSI 1527
Db 3 ECATEDSFV-DDPEEPTIKLNFEFTQNLKSYMQANNIEEDADMSKALVAITPEVASI 61
QY 1528 PTPKLNISRLRTERHQVYELPDSHRLDGMKRPDPDPSYLLAIWTGCTANSQAQPEQ 1587
Db 62 PTPKLNISRLRTERHQVYELPDSHRLDGMKRPDPDPSYLLAIWTGCTANSQAQPEQ 121
QY 1588 KCGKASGKMCDFETCSECNLSREANSQTVRGTLIPCRMTARGSPFLNGTYFQVNELEFA 1647
Db 122 VCNSENGELCASCTFCSCNISREASQAQKVRGTLIPCRMTARGSPFLNGTYFQVNEVEFA 181
QY 1648 DHESLKPIDVPRDWIWLPRRTVYFGTSTVTSIFRGLSTEQIQFCFWKGVCVRGFEQK 1707
Db 182 DHDSRRNPIDVPRDWIWLPRRTVYFGTSTVTSIFRGLSTEQIQFCFWKGVCVRGFEQK 241
QY 1708 RAPRLMARLHPASKLNKNT 1728
Db 242 RAPRLMARLHPASKLNKNT 262

RESULT 7
Q49498 PRELIMINARY; PRT; 917 AA.
ID O49498
AC O49498
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
```

```
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE HYPOTHEICAL 106.3 KDA PROTEIN.
GN F28A23.180 OR AT4G34060.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Bevan M., Weichselgartner M., Fartmann B., Granderath K., Dauner D.,
RA Herzl A., Neumann S., Hoheisel J., Jesse T., Heijnen L., Vos P.,
RA Mewes H.W., Mayer K., Schueller C.;
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Weichselgartner M., Fartmann B., Granderath K., Dauner D., Herzl A.,
RA Neumann S., Mewes H.W., Lemcke K., Mayer K.F.X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL021961; CAAL17566.1; -.
DR EMBL; AL161584; CAB80123.1; -.
DR InterPro; IPR003265; Endo.3c.
DR InterPro; IPR003651; Fes_bind.
DR Pfam; PF00730; Hhh-GPD; 1.
DR SMART; SM00478; ENDO3C; 1.
DR SMART; SM00525; FES; 1.
KW Hypothetical protein.
SQ SEQUENCE 917 AA; 106310 MW; 302B3770ACE1A454 CRC64;

Query Match 7.8%; Score 709; DB 10; Length 917;
Best Local Similarity 21.4%; Pred. No. 3.5e-37;
Matches 232; Conservative 108; Mismatches 163; Indels 580; Gaps 22;

QY 651 IAEIIRYMQNLVGLGDEREQEQNAWLVKGGALVPYESKKRPRKVDIDDETRTNWL 710
Db 388 IAKIKDMGRKLKINK-----VTTMIKADKKLV-----TAKNLDPEIKEMDV 431
QY 711 LMKGDEKEGDEKDKKKKWEERVRFRGRADSFARMHLVQGRFRFPWKGSVVDSV 770
Db 432 LM--VNDSPRSYDDKTEAKKKEREIFQTRIDLFINRMHRLQGNKFKQWKGSVVDSV 489
QY 771 IGVLTONVSDHLSSSAFMSLAARPPPKLSSSRDERNVRSVVVEDPEGCILNLNEIPSW 830
Db 490 VGVFLTQNTDYLLSSNAFMSVAARFP-----VDAREGLSYITEEPQ----- 530
QY 831 QEKVQHPSDMEVSGVDSGSKQLRDCNSGIERNFLEKSTQNLEEVLSQDSFDPF 890
Db 531 -----DAKSECI----- 538
QY 891 QSGRGVSGSCSKSDAEFPTRCTKTSGTSGVTSQVQSGPNLSDICLOGNERPHLYEGS 950
Db 539 -----ILSDE----- 543
QY 951 GDVQKQETTVAOKKPDLEKTMNWKDVCFGQPRNDTNWQTPSSSYEQCATRPHVLDI 1010
Db 544 ----- 543
QY 1011 EDFGMQEGELGYSMWSTSPRVDRVKNKVPFRFRQGSVPREFTGQIIPSTPHELPGMG 1070
Db 544 ----- 543
QY 1071 LSGSSSAVQEHQDDTQHNQOENMKASHLQKTLFDLLNSSECITROSSTKQNTDGLCP 1130
Db 544 ---SISKVEDHEN-----TAKRKEKGTII-- 565
QY 1131 RDRTAEDVDPPLSNSSQLNILVESNSSNKEQTAVEYKETNATILREMKGTLDGKKPTS 1190
Db 1131 ----- 1190
```

```
Db 566 -----EDEIVD----- 571
Qy 1191 QWDSLRKDVNEGROERNKNMDSIDYEAIRRASISEAISKERGMNNLAVRIKDFL 1250
Db 572 -WNLR- MYTKEG--SRPEMHMSVNWSDVRLSGQNVLETIKRGQFRILS----- 620
Qy 1251 ERIVKHGGIDLEWLRESPPDKAKDYLLSIRGLGLKSVCEVRLTLHLNLAFFVDNTNVGRI 1310
Db 621 ERIL-----VDNTNVGRI 632
Qy 1311 AVRGMVPLQPLPSLQLHLELPVLESIOKFLWPLRCLKLDQRTLYELHYQLITFGKVF 1370
Db 633 AVRGLVPLEPLNGVQMH-----QLFEYELHYQMITFGKVF 669
Qy 1371 CTKSRPNCNACPMRGECRHFASAVASARLALPAPEERSLTSATIPVPESP-----PPVA 1425
Db 670 CTKTIPNCNACPMKSECKYFASAYSVSKVLLESPEEK-----HEPNTFMAHSQDVA 722
Qy 1426 IPMIELPLPLEKSLASGAPSNRENCEPIIEEPASPGOECTEITESDIEDAY-----NEDPD 1482
Db 723 VDMTSINLVECVSSGCSQDAICYKPLVEPPSPRAEIPEST--DIEDVPMNLYQSYA 780
Qy 1483 EIPIKLNIEQFGMTLREHM--ERNMELQEGDMSKALVALHPTTITSIPT---KLKNLSR 1537
Db 781 SVPKIDFDLALKSKSDALVISGRMSSDEEISKALVIPTPENACIPKPPRKMYYNR 840
Qy 1538 LRTEHOVYELPDHRLDGMKREPDPSPYLLAIWTPGETANSAPPEQKCGGKASKM 1597
Db 841 LRTEHVYVLPDNEHLLHDFERKLDPSYLLAIWQP----- 878
Qy 1598 CFDETCSECSNLRANSQTVRGTLLIPORTAMRGSFPLNGTYFQVNFELFADHESLAPID 1657
Db 879 ----- 878
Qy 1658 VPRDWIDLPRRTYVFGTSVTSIPRGLSTEIQFCFWKGFVCVGRGFEQKTRAPRPLMARL 1717
Db 879 -----GFLCLRAFDRKQDRKELVRRL 900
Qy 1718 HFP 1720
Db 901 HTP 903

RESULT 8
Q94LX6 PRELIMINARY; PRT; 119 AA.
AC Q94LX6.
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DE PUTATIVE FPPSYNTHASE1 (FRAGMENT).
GN ETPPSSI.
OS Euphorbia tirucalli.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids I; Malpighiales; Euphorbiaceae; Euphorbia.
OX NCBI_TaxID=142860;
RN [1]
RP SEQUENCE FROM N.A.
RA Asada T., Fukusaki E., Kobayashi A., Sando T.;
RT "Prenyltransferase of Euphorbia tirucalli.";
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB051105; BAB47587.1; -.
FT NON_TER 1
FT NON_TER 119
SQ SEQUENCE 119 AA; 13771 MW; FDBDCE4CD0CC5B01 CRC64;

Query Match 5.0%; Score 458; DB 10; Length 119;
Best Local Similarity 78.7%; Pred. No. 2.8e-22;
Matches 85; Conservative 10; Indels 0; Gaps 0;

Qy 1214 DSIDYEAIRRASISEAISKERGMNNLAVRIKDFLERIVKHGGIDLEWLRESPPDKA 1273
```

```
Db 12 DYLDWEAVRCADVDEIANTIKERGMNNVLAKRIKDFLNRLVREHGNIDLEWLDPVDPKA 71
Qy 1274 KDYLSSIRGLGLKSVCEVRLTLHLNLAFFVDNTNVGRIAVRMGWVPLQP 1321
Db 72 KEYLLSIRGLGLKSVCEVRLTLHLNLAFFVDNTNVGRIAVRMGWVPLQP 119

RESULT 9
Q9NG79 PRELIMINARY; PRT; 2151 AA.
AC Q9NG79.
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DE HYPOTHETICAL 232.4 KDA PROTEIN.
OS Trichomonas vaginalis.
OC Eukaryota; Parabasalidea; Trichomonadida; Trichomonas.
OX NCBI_TaxID=5722;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=#202;
RA Delgaty K.L.;
RT "The Molecular Identification and Characterisation of a Candidate Gene
for Cell Detaching Factor from Trichomonas vaginalis.";
RL Thesis (2000), University of Ottawa, Ottawa, ON, Canada, In press.
DR EMBL; AF257323; AAP67197.1; -.
KW Hypothetical protein.
SQ SEQUENCE 2151 AA; 232366 MW; F2EF91C5D9C60C82 CRC64;

Query Match 3.1%; Score 281; DB 5; Length 2151;
Best Local Similarity 19.7%; Pred. No. 5.6e-09;
Matches 367; Conservative 256; Mismatches 705; Indels 536; Gaps 88;

Qy 10 VNATEATEQNDGSRQDVLEFDLNTPOOKPSRKRKPMKVVEG-----KPKRK 59
Db 372 LNLKPSQSSQTSSEKPKLSNLGKSPSNQSTEEK---PKLQGGIKLGQTSQNKSEK 427
Qy 60 PRKPAELPKVVVEGPKP-----RKPKAATQEVKSKETGSAKKKNLKESATKP--- 109
Db 428 PKLSNLGKSPSTEEKPKLSNLGKSPSNQSTEEKP---LGFNLPKAPSNQTEEPKLG 484
Qy 110 ----ANVGD--MSNKSPEVTILKCRKALNFLENPGDARGDSESEIVQNSGANSFSEI 163
Db 485 GGISLNLGNKPSQSEKPKLSLGGIKLA-----QSPSNSNEEKPKLSNLKPSNSQSTEE 539
Qy 164 RDAITGNTGFLDSVQIDKTNGLGAMNQPLEVSMGNQPD-----KLSTGA-KLARDQQP 217
Db 540 KPKLGFNLPKAPSNQTEEPKLGTTGG---ISLNLGNKPSQSEKPKLSLGGIKLA--QSP 593
Qy 218 DLLTRNQQCFPVATQNTQPMENQOAWLQMKNOLIGFP--FGNQQPMRTIRNOOPCLAM 275
Db 594 S----NSNEEKPKLSNLKPSNSQSTEEKPKLQGGIKLNLGNKPKQTEETEEKPKLQL 649
Qy 276 G-----NQPMYLLIGTPRALVSGNQQLGG-----PQGNKRPIFLNHTCLPAGNQLY 323
Db 650 GGIKLGQSPSNST--EEKPKL-----QLGGIKLNLGSKPQTEEPKPL-----QLG 692
Qy 324 GSPTDMHQLVMSTGG-----QOHGLLNKQPGSLIRGOQPCVPLIDQ 366
Db 693 G-----IKLGTGGISLNLGNKPSQSEKPKLQGGIKLGNQSP-----NQ 731
Qy 367 QPATPKGFTLHNQMVATSMSSPGLRPHSQVPTTYLHVSVSRILNGTTGTTCORSRAPA 426
Db 732 PLEKPSGINLN-----LGKSQP-----SSEKPKLGLN-----GKSPSNSTEEPK 774
Qy 427 YDSLQQDIHQGNKYILSHEISNGCKKALPQNSSLPTIPMAKLEEARSKROYHRAMGQ 486
Db 775 LGTGGISLNLGNK-----PQTEEP----- 794
Qy 487 TEKHDNLNLAQOIA--OSQDVERHNSSTCVYLDAAKTKIO-KVVOENLHGMPPVEIED 544
```

Db 795 --KLSLNLKSPSNQONQSTE-----EKPQLQLGGLKLNGLNKPQTETQTE 838
Qy 545 DP---TDGARKGNATASISK-----GASKGNSSPVAKTAKEKCIVPKTPAKKGRAGR 594
Db 839 KPKQLGGLKLNGLSKSQTEKPKFQLGGIKLQGSQSNSTEKPKLSL-----GG 888
Qy 595 KSVPPPHASEIQWQTPPKTPLSR---SKPK-----GKRKS-----IQDSG 636
Db 889 IKLAQSPNEEKPKL--SPNLKPSNQOOTEKPKLQLGGLKLNGLSKSQTEKPKLSSEGG 947
Qy 637 KARGPSGELLQODSTAETIIRYONLYLGDKEKEQONAM-----VLKGDGALVPY 687
Db 948 IKLGNVSSQTSDEKPKLUGLGGISFNQOOTEKPKLSSQNVKPTLGLGGITLQOQ 1007
Qy 688 ESKRKPR-----PKVDIDDETRTNWLLMGK-----DEKEGDEKDK----- 726
Db 1008 TSEENKPKLGNLPSNQTEKPK-----LGTGGISLNLGNKPQSEKPKLQLGGIKLAH 1062
Qy 727 KKEKWEEERRVFRGRADSFARMHVLQDRRFSPWKGSVVDSS-VIGV-FLTQNVSDHLS 784
Db 1063 SPSNQSTEEK-----PKLSNLGK---SPSNQSTEEKPKLGLGGLKLNGLNHQ 1108
Qy 785 SSAFMSLAARPPKLSRSSREDERNVRSVVVDEPECIILNLAIEPSQW-EKVQHPSDMEVS 843
Db 1109 SDE-----KPFSLGCMKLGKVPNESQEPKFTLNIKPLFSNQAEQNDGPKLGLG 1160
Qy 844 GVD---SGSKQLRDCNSGI-----ERFNFLEK-----SIQMLEEVLSS----- 881
Db 1161 GISLNLQSKKEENEKPKLGIAPKPIQSNQNKIEKPKVTNSVTKTEKAISAPGVKLDL 1220
Qy 882 ---QDSFDPALFOSCGRVGSCSKSDAEPPTTRCETKTVGSGVSQSVQGTGSP 930
Db 1221 KLPKKGFLMRQPSKSLSEVSNDSQNLSSFGQIHVTLPEKKEENQPLSSFGQ-IHVTL 1279
Qy 931 NLSDEICLQGNRPHLYEGSGDV-----OKOETTNVAOKKPDLEKTMNWKDVCFGQ--- 982
Db 1280 EKT-----QKRENOPLSSFGQIHTLPQKQNEKVFQKVTDEKLS-----SFGQIHV 1329
Qy 983 ---PRNDTNQ---TTSSSYEQCATRPHVLDIEDFGMQE---GLGYSWMSISPRVDRYKN 1036
Db 1330 TIPEKLANQOEKPLSTPEQI-----HVSIEKTKQOEKISSFGQIHVSIPEKLOKQEE 1384
Qy 1037 K-----NVPRRFFRQGG-----SVPREFTQGIIPSTPHEL-----PG 1068
Db 1385 KTSFEGQIHVTIPEKLOKQEGSVIDKLP-EKTOELIKNMPESEINVKTIPEQVHQSQSNPL 1443
Qy 1069 MGLSGSSSAVOEHODDTQHN-----QODEMKNASHLQKTFDLNLN-----SSECLT 1115
Db 1444 VPLGVSLNDVAKSQNTQNKPDLSLKPQKSEENKPS-----LGLKLPNKNFNSLKPQ 1499
Qy 1116 ROSSTKQNTDGLPRDRTAEDVDPLSNSSLO-NILVSNSSNKEQTAVEYKRETNATI 1174
Db 1500 SQTESKPNLSSLLPPKQTL-GLKLPTSSNGLKLPKLPQNQRNNOEAKVEKET---- 1554
Qy 1175 LREMGTLADGKKPTSQWDSLRKDVEGNEGRQE-----RNKNMDSIDYEATRASSIS 1230
Db 1555 -----KEYKAKEPKKELGFKSDFKIEGASPKGINLSSSGKF- 1591
Qy 1231 EAIKRGHNNMLAVRIKDFLERIVKDHGGIDLEWLRSPDPKA-KDYLL-----SI 1280
Db 1592 EYPSGKSFDDMIKQVNAEENKPKFSNLNLNKNKTPLSNAPKPIALSISNCGKPKSL 1651
Qy 1281 RGLGLKSVCEVRLTLHLNLAFPVDTNVGRIAVRMGWVPLQPLPESLQLHLLLEYPVLES 1340
Db 1652 AGLNISKSSLASVSINALDSGD-DVPEKLSIDKLKKEPPPP-----PTPAV 1700
Qy 1341 OKFLWPRLC---KLDQRTLYELHQLITFGVCTKSRPNCNAC-PMRGECHRFASAYAS 1396
Db 1701 TK---PALATEKEIRESSIADV-TPLAMITKINSK-KPDFSALKPNLGNKQSSSSNEQ 1755
Qy 1397 ARLALPAPEERSLTSATIPVPPEFPPVPAIPMIELPLPLEKSLASG-----APSNRENC 1450
Db 1756 KPLANKPNFSALKLPQKSEETKPNLSSLLPPKQTLGLKLPKLTSSNSGKLPKLPQNQNOQ 1815

Qy 1451 EPIIEE-----PASPQOECTEITESDIEDAYYN-----EDPD----- 1482
Db 1816 NOEVKESKEVKAKEPKKELGFKSDFKIEGASFKGINLSSSKKEFVPSGKSFDDMIKOKA 1875
Qy 1483 --EPTTIKNTIEQGMTLREHMERNMELQ-----EGDMSKALVALHPTTTSIPTP 1530
Db 1876 KTEQPKTEBNEPKFSFNLNLNKNKTPLSNAPKPIALSLSNDSNKGQVQGALTIKPIP-- 1933
Qy 1531 KLNISRLRTEHQYVELPDSHRLLDGMDKREPDDOPSPVLLAIWTPTGETANSQOPEQKCG 1590
Db 1934 -LSKGDKPKADFQSLKLGKS---LSNPTQNKPD-----LSSLKPPQKSEE 1974
Qy 1591 GKAS 1594
Db 1975 NKPS 1978

RESULT 10
O76891
ID O76891 PRELIMINARY; PRT; 5327 AA.
AC O76891;
DT 01-NOV-1998 (TReMBLrel. 08, Created)
DT 01-NOV-1998 (TReMBLrel. 08, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE EG:49B4.1 PROTEIN.
GN FUTSCH OR EG:49B4.1 OR CG3064.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephyroidae; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RA Papagiannakis G., Spanos L., Siden-Kiamos I., Louls C.;
RT "Sequencing the distal X chromosome of Drosophila melanogaster.";
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Benos P.;
RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
RL EMBL; AL031128; CA20006.1; -
DR Flybase; FBgn0015390; futsch.
SQ SEQUENCE 5327 AA; 575942 MW; FEFEE23A118FF38A CRC64;

Query Match 3.1%; Score 278; DB 5; Length 5327;
Best Local Similarity 19.7%; Pred. No. 3.4e-08;

Matches 277; Conservative 198; Mismatches 583; Indels 348; Gaps 59;

Qy 2 QSIMDSSAVNATEATEQNDGSRQDVLEFDLNTKTPQOKPSKRKRKMPKVYVVEG-----K 55
Db 1986 ESID-----EAEKSESRRE-----SVAEKSPLPKSEASRPASVAESIKDEAEK 2031
Qy 56 PKRPRKPAELPKVYVEGPKRKP-----RKAATQKVKSKETGSAKKNLKESATKK 108
Db 2032 SKEESRRESVAEKSPLPKSEASRPASVAESIKDEAKSEESRRESAAEKSPLPKSEASR 2091
Qy 109 PANYG-----DMSNKSPEVTLKSCRKALNFLENPDGAR--QGDSE--SEIVQNSGANSF 160
Db 2092 PASVAESVKDEADSKSESRRE-----SMAESGKAQSIKGDQSPLESKSRPESV 2144
Qy 161 SEIRDAIGTNGSFLDSVSQIDKTNGLCAMNQPLEVSMGNQPKLSTGAKLARQOQPDLL 220
Db 2145 KD--DPVKSKEPSRRESVAGSVTADSDQSPLESKASRPESVDSVKDEAKQESRR 2202
Qy 221 TRNQCOFFPVATQTFPMENQQAWMKQLIGFPFGNQQRMTIRNQOCLAMGNQOP 280
Db 2203 ESKTESVIPPKAKDKSPKEVLQ-----PVSMTPE---TIREDA-----DQP 2240
Qy 281 MYLGTFRPALVSNQO-----LGGQGNKRPFIPLNHQTLCPAGNQLYSGPTDMHOLV 333
Db 2241 M-----KPSQAESRRRESIAESIKASSPRDEKSPKSLASKEASRPGSVAESIKYDLDKPOII 2294

```
QY 334 MSTGQQH-----GLLIKN-----QQGSLIRQQPCVPLID--QQPATPK-----GF 374
Db 2295 KDKSTEHSRESLEDKSAVTSKSVSRPLSVASDHEAAVAIEDDAKSSISPKDKSRPGF 2354
QY 375 THLNQWVATSSPGLRPHSOSQVPTTLVHVESVR-----ILNGTGTQCRSRAPAYD 428
Db 2355 -----VAETVSSP-----IEEATMEFSKIEVVEKSSLALSLOGSGGKLQTDSSPV-- 2400
QY 429 SIQQDHQGNKYLHSHEISNGNGCKKALPONSSLPPTMAKLEEARGSKROYHRAMGOTE 488
Db 2401 ----DVAEGD---FSAVA-----SVSTVPTLTTPAELA-----QIGAATV 2436
QY 489 KHDLLAQIAQS--QDVERHNSSTCYEYLDAAKTKIOKVVOENLHGMPEVETIEDD-- 545
Db 2437 SSPLDALRTSPAPHISRA DSPACASEEIASQDKSPQVLKES--SRPANVAESKDAA 2494
QY 546 ----PTDGARKGNKTASTISKASKGN--SSPVKKT-----AEKEKIVPKTPAKKRA 592
Db 2495 QLKSSVEDLRSPVASTEISRPASAGETASPIEEAPKDFAEFEQAEKAVLPLTIELKGNL 2554
QY 593 GRKKSVPPAHASEIQLMQPTPPKPTPLSRKPKGKRSIQDSGKARGPSGELLCOQDSIA 652
Db 2555 PTLSSPVDVAHAS--VQPAELSKVDIEKTASSPIDEAPKSLIGSPAERPERPSAPSAKDA 2613
QY 653 EIIYRMQNLXLYLGDKEREQEQNAMVL-----YKGDGALVPYESKKRPRPKVDIDDETT 705
Db 2614 ESV-----EKSKDASRPSPVVESTKADSTKGDISPSP--ESVLEGPDKDDVEKSESS 2663
QY 706 RIWLLMGKEGEKEDEBKDKKKEKWEERVRFRGRADSFIA RMHLVQGDRRFSPWKG 765
Db 2664 RPPSV-----SASITGDSTKDVSR-----PAS 2685
QY 766 VYDSVIGVFLQNVSDHLLSSAFNSLAARFPKLLSSREDERNVRSVVVEDEPGCILN 825
Db 2686 VVESV-----KDEH-----DKAESRRESIAKVESVI--DEAG----- 2715
QY 826 ETPSQEVRVQHPDSMEVSGVDSGSKQELRDCSNSGIERFNEFKSIQNL-----EVL 879
Db 2716 -----KSDSKSSQDSQDEKSTLASKEASRESVSVESKDDAEKSESPPESVI 2764
QY 880 SS-----QDSFDPAIFQCGRGVSGC-----SCSKSDAEFPPTTRCETKTVSGTS 922
Db 2765 ASGEVPVRESPLDSKDTSRPGSVVESVTADEKSEQQSRRESVAESVKADTRK--DGKS 2823
QY 923 OSVOTGSPNLDEICLOGNE---RPHLYEGSGDVQKQETINVAQKPKDLEKTMNWKDSV 978
Db 2824 Q--EASRFSVDELKDDDEKQESRRQITGSHKAMSTMWGDESMDKADKSKESRSPESV 2881
QY 979 CFGQPRNDTNQTTTPSSSYEQCATRQPHVLD--IEDFGMQEGGLGYSMMSIS--PR--VDRV 1034
Db 2882 AESIKHENTKDEESPLGS-----RRDSVAESIKSDITKGEKSPLPKSREVSRRPESVSGSI 2935
QY 1035 KNKNVPRFRFGGSGVPREFTGQII---PSTPHELPGMGLSGSSSAVOEHQDDTQHNOOD 1091
Db 2936 KDEAESRRESVAESVKPESKSDATSAPPKSEHSRP-----ESVLGSLKD 2980
QY 1092 EMNKASHLOKTFDLNLSSECLRQSTK-----QNTDGCCLPRDRTA-----EDVVD 1140
Db 2981 EGDKTTSRVSVAGSIKDEKSLIVSQEASRPERESAESLUKDAAPSQETSRPESVETESVKD 3040
QY 1141 PLSNSSQLNVLVESNSSNKQETAVYEKTNATILREMK--GTLADGKKPTQOWDSLRKDV 1199
Db 3041 GKSPVASKEASRPASVAENAKDSADEKQRPESLPQSKAGSIKDEKSPASKDEAEKSK 3100
QY 1200 ENEGRQERNKNNMDSIDYEAIRASISEI--SEAIKERGNMML-----AVRIK 1247
Db 3101 E--ESRRESVAEQFPVLSKEYSRFASVAESVKDEAEKSEESPLMSKDSRSPASVAGSVK 3158
QY 1248 DFLERIVKHGIDLEWRLESPPDKA 1273
Db 3159 DEAEKSE-----ESRRESVAESK 3177
```

```
RESULT 11
Q95YM2
ID Q95YM2 PRELIMINARY; PRT: 17352 AA.
AC Q95YM2;
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE I-CONNECTIN.
GN I-CON.
OS Procamburus clarkii (Red swamp crayfish).
OC Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;
OC Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Astacidea;
OC Astacoidae; Cambaridae; Procamburus.
OC NCBI_TaxID=6728;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21423462; PubMed=11532946;
RA Fukuzawa A., Shimamura J., Takemori S., Kanzawa N., Yamaguchi M.,
RA Sun P., Maruyama K., Kimura S.;
RT "Invertebrate connectin spans as much as 3.5 micrometer in the giant
RT sarcomeres of crayfish claw muscle.";
RL EMBO J. 20:4826-4835(2001).
DR EMBL; AB055861; BAB64297.1; -.
SQ SEQUENCE 17352 AA; 1962348 MW; 4BA157BEC042E42D CRC64;

Query Match 2.8%; Score 253; DB 5; Length 17352;
Best local Similarity 17.9%; Pred. No. 7.9e-06;
Matches 364; Conservative 336; Mismatches 789; Indels 550; Gaps 88;

QY 6 DSSAVNATEATEQNDGSR-----OD---VLEFDLNKTPQ---OKPSKRK 43
Db 11703 ESISVSQTOVHESVDDVRVAKLBEKALTILPQDSITVQELSVKEAPQIDKHKPKTEQ 11762
QY 44 RKFM-----PKVVVEGKPKRKRKP-----AELPKVVV 71
Db 11763 AKSILSPRESTAVEINEKPSVSDVVKPTEQA KSIIPSHESLTVEVTVKETPKDIS 11822
QY 72 ECKPKRPRKAATOKYKSKETGSAKKNLKESATKRPANVGDMSNKSPET--LKSCRKA 130
Db 11823 EVKPKQTA KSIWSE---QESISVQEVTVKDA---PGSVKDDKPKTETATSIIPSHES 11874
QY 131 LNFD---LENPGARQGDSESEIV-----QNSGANSFSEIRDAIGGTNGSFIDSVSQ 180
Db 11875 LTQEVSVKAEAGDIDKDKPKAEERATSI LSPQESITAVEVS--VREA---PGSVKDSIPK 11929
QY 181 IDKTNGLGMNQPL---EVSNGNQPKLS-----TGAKLARD 214
Db 11930 KEQATSIISTHESVTVEHVTKPEPKDISDKPKTEKATSI LSEQESISVEEVSVEAPG 11989
QY 215 QPDLLTRNQOCQFPVATQNTQFPMEHQOAMLOMKNOLIGFPFGNQOPRMTIRNOOPCLA 274
Db 11990 SVKDLKUTEQAASIIIPSHES---LTQEVTVKEAPVEITDKKPKKEKATSI LSEQESIA 12046
QY 275 MGN---QOPMYLIGTPRP-----ALVSGNQOL-----GPOQNKRP-----IF 309
Db 12047 VQEVSVKEAPYSMEEDPKAEKATSIIPSHQSLTIEEVTVEKASGNI SDKPKTEKALSI 12106
QY 310 LNHOTCLPAGNOLYGSPTDMHQLVNSTGGQOHLIKNQPGSLIRGOQPCV----- 361
Db 12107 LSEQESISV-----QEVSVKEAPGSVKETKPKTEQATSIIPSHESLTVEHVSVE 12156
QY 362 --PLIDQOPATPKGFTHLNQMVATM-----SSPG---LRHQSQOV---PTTVLH 404
Db 12157 AHQDLSQDKLTERATSVLSEQESISIQEILVKDAPGSVKDIPKPKSEQATSIIPSHESLT 12216
QY 405 VESVSRILNGTTGTCQSRAPAYDSLQODIHQGNKYILHSHEIS--NGNGCKKALPQNSSL 462
Db 12217 VQEVSVKESPSEISDKRPKSEKATSI LSE-----QESISVHESVVDAPGSKMDAKPKTEQ 12272
QY 463 PTPIMAKLE-----EARGSKROYHRAMGOTEKHDLNLAQIAQS--QDVERHNSSTCVC 514
```

```
Db 12273 ATSVISPHESLTVQEVSVKSSSTSEISDKPKSEKATSIILSEQESISVQEIISVDKAPGSIK 12332
QY 515 YLDAAKTKTKQVQENLHGMPPEVIEEDPTDGARKGNTASISKAGSKGNSSPVKKT 574
Db 12333 --DAKPKTEQATSIISPHESLTVQEVSVKETPTDISDRKPKTEKATSIILSEQESISVQEV 12390
QY 575 AKEKCKIVPWT-PAKGRAGRKSVPPPAHASEIQLWQPTTPKTPLSRSPKPKGKRKSI- 632
Db 12391 SVKE---APGTVETAKPKTEQATSIISPHESISVQEIISVKEAPTDICDKPKPKSEKATSIL 12447
QY 633 --ODS-----CKARG-----PSGELLCDQ-----SIAEIIYRMQNLYLGKERE 669
Db 12448 SEQESIAVQEVSVKEAPGSEIEAKPKTEYAKSSISPHESLTVQEVVLVKEASAEISDKRPQ 12507
QY 670 QEQNAMLYLKGDALVYESKKRPKVDIDDETTIWNLLMG-----KGDEK 718
Db 12508 TGRKATSIILSEQESIAVQEVSVKEAPGRVEDVKPKTEQATSIILSEQESLTVQEVTVKGD 12567
QY 719 EGDEKDKKKEKWEERRVFRGRADSFIAHMLVQ-----GDRRFPWKGS 765
Db 12568 EISDKPKTE-----KAISILSEQESISVQEIITVKEAPGSKVEDKLPKQAT 12614
QY 766 VVDSVIGVFLTONVSDHLLSSAFMSLAARP--PKLSS--SREDERNVRSVVVEDPGEI 821
Db 12615 SVISPLESLTVQEVSVKESG---DISDRPKTEKATSIIFSEQESISVQEVSVKEAPGT 12671
QY 822 LNLNEIPSWOEKQHPDMEVSGVDGS-KEQLRDCNSGTERFNFLEKSTQNLEEVLS 880
Db 12672 ADVK--PKEQATSIILSEQESLTVQEVTVKEIPTDISD-----KKPKSEKATSIILS 12720
QY 881 QDSFDDPAIFQSCRGVSGSCSK-----SDAEFTTTCRETIVS 919
Db 12721 EQESISVHEVSVKDPAGSMKDAKPKTEQATSVISPHESLTVQEVSVREVPTSEISDKPKS 12780
QY 920 GTSQSVOTGSPNLS-DEICLQGNRPHLYEGSGDVQOQET----- 958
Db 12781 EKATSIILSEQESISVQEVSVK--EAP-----GSMKDAKQKTEQATSVISPHESLTVQEVSV 12834
QY 959 ----TNAQAQKPDLEKTMNNKDSVCFQPRNDTNMQWTPSSSVBQCATROPHVLD----I 1010
Db 12835 KEAPTELSDRKPKEKA-----TSILSEQESISVQEVSVKEAPGCV 12875
QY 1011 EDFQMGQEGGLYSWMSISPR-----VDRVKNKNVPRF-----FRQGSVPREFTGOIIP 1060
Db 12876 KDLKLTKE---QATSVISPHESLTVQEVSVKEAPTEISDRKPKTEQATSIILSEQESISI- 12931
QY 1061 STPHELPGMGLSGS-----SSAVQEH-----QDQTOHQQQDEMKNKASHLQKTF 1104
Db 12932 ---QEIISVKGAPGLKDEKPKSEQATSIILSEQESLTVQEVTVKEAPADISDVKPKTEKAT 12988
QY 1105 DLLNSSECLTRQSSSTKO---NITDGLPDRDRTAEDVVDPLSNSSNQNLV----- 1153
Db 12989 SILSEQESISVQEVSVKEAPGSMKD-AKPKTEQATSVISP-HESLTVQEVSVKEVPTEIS 13046
QY 1154 -----ESNSSNKEQTAVEYKETNATILRMKGTADGKKPTSQWDSL----- 1195
Db 13047 DKPKSEKATSIILSEQESISVHEVS---VKESPGSMKDAKPKTEQATSVISPHESLTVQ 13103
QY 1196 -----RKDVEGNEGROGRKNNWMSIDYAIRASTISEIASEIKE-RGNMNLAVRIKD 1248
Db 13104 VSVKESPTESLDRKPKKEKATSIILSEQESISVQEVSVKEVPESMKDLKPKTEQATSVISE 13163
QY 1249 FLERIVKHGGIDLEWLRESPPD-----KAKDVLISIRGLGLKSVCEVRLTLHLNL 1299
Db 13164 LLSUSVQEVSV-----VKESPGSLSDKKPKTERATSIILSEQESISVQEVSI- 13207
QY 1300 AFPVDTNVGRIAVRMGVPLQPLPESLQLHLELLEYPVLESTQKPLWPRCLCKLDORTLYEL 1359
Db 13208 -----VKAPGSVE---EVKPKTEQATSIISPHM-----SL 13235
QY 1360 HYQLITGKVKFCYSK---RPNC-NACPMRGECRHFA-----SAYASARIALPAPEERSL 1409
Db 13236 TVQEVTVKEVPADISDVKPKSEKATSIILSEQESISVQEVSVKEAPGSMKDAKPKTEQ--- 13292
```

```
QY 1410 TSATIPVPPSPFPVAIPMIELPLPLEKSLASGAPSNRENCEPIIEFPASPQGECTEITE 1469
Db 13293 -ATSVISPHESLTVQEVSVREVPTETI-----SDKKPKSEKATSIILSE-----QESISVQE 13341
QY 1470 SDIEDAYNEDDPEIPIKUNIQFGMTLREHMERNMELQEGDMSKALVALHPTTTTSIPT 1529
Db 13342 ISVKEA-----PESMKDAKPKTEQATSVISPH-----ESLTVQEVSVKEV-----PTEISDKK 13389
QY 1530 PKLKNISRLRTEHQ-----VVELPDSHRLLDGMDKREPDPSPY-----LLAIWT 1574
Db 13390 PKYKATSIILSEQESISVQEVSVKEAPGSKVDLKPKEQATSVISPHESLTVQEVSVKEA 13449
QY 1575 PGETANSQAQPEQKCGGKAGKMGCFDETCSECNSREANSQTVRGTLILPCRTAMRGSFP 1634
Db 13450 PTEISDKPKPKTEQ-----ATSVLSEQESI-----SIQESISVKEAPGSL-----KDEKP 13492
QY 1635 LNGTYFOVNFELFADHES-SLKPTDV---PRDWLWDLPRRTVYFGTSVTSIFRGLSTQOI 1689
Db 13493 KSE---QATSIILSEQESLTVQEVTVKEAPAD-ISDVKPKTEK-ATSIILSEQESISVQEI 13546
RESULT 12
Q9HCY0 PRELIMINARY; PRT; 2254 AA.
AC Q9HCY0;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE PUTATIVE TRANSCRIPTION FACTOR-LIKE NUCLEAR REGULATOR.
GN TFNR.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21100861; PubMed=11161782;
RA Kelter A.R., Herchenbach J., Wirth B.;
RT "The transcription factor-like nuclear regulator (TFNR) contains a
RT novel 55-amino-acid motif repeated nine times and maps closely to
RT SMN1."
RL Genomics 70:315-326(2000).
DR EMBL; AJ238520; CAC04245.1; -.
DR InterPro; IPR000977; DNA_ligase.
DR InterPro; IPR001005; Myb_DNA_bind.
DR Pfam; PF00249; myb_DNA-binding; 1.
DR SMART; SM00395; SANT; 1.
DR PROSITE; PS00697; DNA_LIGASE_A1; UNKNOWN_1.
SQ SEQUENCE 2254 AA; 252844 MW; F350E96F53F04CFE CRC64;
Query Match 2.7%; Score 245; DB 4; Length 2254;
Best Local Similarity 18.7%; Pred. No. 1.3e-06;
Matches 358; Conservative 250; Mismatches 722; Indels 584; Gaps 85;
QY 11 NATEATEQNGDSROD-----VLEFDLNTPOOK----PSKRRKRKMPKVWVGKPKRK 59
Db 339 NKFKREKTNWRIDKAFQEKRPDFDFFAHLLQKVLAEKREKQKQKVNHSLEKSKTK 398
QY 60 PRKPAELPKVVVEG-----KPKRKPKRAAT----- 84
Db 399 PRKNVKKVACGCVNNDPDESSSRISDTERSOKDAQTVEEESLTLSDREDAQVALEVD 458
QY 85 --QEKVSKETGSAAK---XNLKESATKK--PANVGDMNK----SPEVTLKSKRK---- 129
Db 459 LNQKRRRKQGDGANELGVNLLNENATVQAGPSKGEKHKNKQAIKQAIKELKEGCKEQML 518
QY 130 --ALNFD-----LENPGRQGDSESEIVQNSS-----GANSFS 161
Db 519 SCTQNDIGIVGFASTEKEVKRTDFILSLSNQDATSVATSESESSTDLPSFEVIGALC 578
QY 162 EIRDAIGG---TNGSFLDSVQIDKTNGL-----GAMNOPLEVSMGNQPKLSTGAKLA 212
```

Db 579 EVNNAEGSCIERVNDLKNNSLEIDOTENVKPMRLGRFORP-----KPNLSRAGKSKV 631
Qy 213 RDQ-OPDLLTRNQOPVATQNTQPMENQOAMLOKKNQLIGFPGNQOQPMTRNQOP 271
Db 632 LSQKTESKNSKSHKSVKKNHVEKKMNTDILRMETTERENP-----EATVSVLGEKN 688
Qy 272 CLAMNQOPMYLIGTPRPALVSGNQQLGGPOGKN-----RPIFLN-----HOTC 315
Db 689 CLQESQ-----LKALRPVQVRGLQPKPNAGRAERKEILISOEEIGANVEKNENESC 743
Qy 316 LPAGNQLYSGTDMHQLVMSTGGQOHLIKNQOPGSLIRQOP-----CVPLIDQ 367
Db 744 A-----DRDTPQHMEDQSRKDEEEDVILQPEKNDSFQVQDEPKVLNECLSVQENN 796
Qy 368 PATPKGFTHLQWATSSSPLGPHSOSQVPTTYLHVESVRLNGTTCQCRSRAPAY 427
Db 797 KA-----NKLQV-----PILTRFQPKP-----NIGRGTRGREISKE 831
Qy 428 DLSQODIHQGNKYILSHBISNGCKKALPONSSLPPIAKLEEARSGSKQYHRAMQOT 487
Db 832 EVLEKILVSGEMAAALRETFLDTPSPKEM-----VPAEINTK-EMQSDLKETGTRAIISPR 885
Qy 488 EKHLNLQAQIAQSQDVERHNSSTCV-----EYLDAAKTKIKQVQOE-----NLHG 534
Db 886 EKILDVIDDITMETGLKAMGREICLREKTPVIDATE--EIDKDEEAGRREISPOKNG 943
Qy 535 MPVEIETEDDPTGARKGKNTASISGASKNSSPVKTKAEKECI--VPKTPAKKGR- 591
Db 944 -PEEVKPLGEVETD-----LKATGNSSPREKTPGEGTDATEEIDKNLEETGRR 990
Qy 592 -----AGRKS-----VPPPAHAS-BIQLWQP-----TPPKTP 618
Db 991 KISPENGPEEVKPVDEMETDLNATGRESSPREKTPVIDATEIDLEETERESVPOENG 1050
Qy 619 LRSKPKGKRKSTQDSGK-----ARGSGELLQCDISAEIYRMQNLYLGKERE--QEQN 673
Db 1051 LEEVKPLGEMETDLKATGROSFPFGKTPPEVI--DAIEEI-----FIDLEETEREISPOEN 1103
Qy 674 AMVLYKGDGAL--VPVESKRRKPRPV-DIDDETTRINLLMGKDEKDEKKKE 729
Db 1104 GLEEVKPLGEMQTLKATGREISPREKTPVIDAT-----EIDKDE 1146
Qy 730 KWEEERVFGRADSPFIARMLVQGRDRFPSPKGSVVDVSVIGVFLFQNVSDHLSASF 789
Db 1147 E-----TGRREISP-----BENGPEEVKPVDEM 1169
Qy 790 SLAARFPKLLSSREDERNVRSVVDEPEGCILNLEIPNSQEKVQHPSDMEVSGVDSGS 849
Db 1170 ETDLKTTGREGSSREKRE--VIDAAEVIETDLE-----TEREISPOENGP 1214
Qy 850 KEQLRDCSNSGIERFNFLEKSIQNLLEEVLSQDSFDPATFQSGRGVSCSCSDAEFP 909
Db 1215 EE-----VKPVGRMETDLKEIREISQRE-----KVLAEFS 1245
Qy 910 TTR-----CET-----KTVSGTSQSV-----QTGSPNL-----SDEICLOGNER 943
Db 1246 AIREKEIDELTKGRDIPIMEKVSGKMAVVEEMADLKETGCKENFRERGSEIICV----- 1300
Qy 944 PHLYEGSDVQKQETTNAVQKKPDLKTMWKNDSVCGQPNNDTNGWTTTSSSYEQCATR 1003
Db 1301 --TEEKVAELKQTKTDISPRENLEETSTSR-----QTDHLMQSGSNDFSAMPS- 1349
Qy 1004 QPHVLDIEDFCMOQEGLYGWSMISPRVDRVKNKNVPRRF--PROGG-----SVPREFTGQ 1057
Db 1350 ----LDLQNTSSE-----VLSMMHTPVEKRNKESEVSHSFHFKISSQTHESDKTEVQGI 1401
Qy 1058 IIPSTPHELPMGLUGSSSAVQEHQDQTQNOQDMNKASHLQKTFDLDLLNSSBECLTRQ 1117
Db 1402 QSPDVPEQFSDINLSKSL-----PQEQKPLEIKPAPFVRSRF----- 1438
Qy 1118 SSTQNTIDCGLPRDRTAED--VVDPLSNSSNLONILVSNSSNKEGTAVEYKETNATIL 1175

Db 1439 KRKPENLARAALKRETTESEKYYIEKKSKTKMETIVQENNEQDTLPSQHDASLMIS 1498
Qy 1176 REMKGTIADGKK-----PTSQW-----DSLKRDVEGNEGROERNKNNDSI 1216
Db 1499 RE-KDTLGHNEEAIVLPCTQTERNLSPNSCEPKEESQAPVQKNDVSVSVGTNNVNTF 1557
Qy 1217 DYEAIRRASISESEAIKERGMNMLAVRIKDFELERIVKDHG-----GIDLEWLRESPPDKA 1273
Db 1558 QQE-----MKSQVQTAQVRGRQLRQPRNIRKQTQORQIVDKGEAKGIKEGRTILPKDET 1613
Qy 1274 KDYLSTIRGLGLKSEVCRLLTLHLNLAFFVDVTNNVGRITAVRWGVPLQPLPESLQHLLEL 1333
Db 1614 EK-----KVLTVSN--SQIETEIE-----VPSSAVPE----- 1638
Qy 1334 YPVLESIQKFLWPLRLCKLDQRTLYELHYQLITFGKVFTKS-----RPNCNACPMRGEC- 1387
Db 1639 -----HRMYENOSQVVLVENLHVNTKNTIRHENKPYVPSSAQMT 1678
Qy 1388 -RHFASAYASARLALPAPEERSLTSATIPVPPESFPVPAIPMIELPLPLEKSLASGAPSN 1446
Db 1679 RRFQKAKPNLGRAHKKKEPVLEKVTDDQSKGKP-----EDHLLQKGSN 1725
Qy 1447 -----RNCNPIIEEPASQOECTEITESDIEDAYYNEDEPDEI--PTIKLNEQFG-----M 1496
Db 1726 TQLLLKEKABELLSLEVSARKDCVGSRESAL--AKIDAEELEVGPSSRRVGEETVGDNSPS 1783
Qy 1497 TLREHMERN-----MELQEGDMSKALVALHPPTTSTPTPKLKNISRLRTEHQVYELPD- 1549
Db 1784 SVVEEQYLNKLTSCPPQPLNETSYSK--IALDGKTTISTSEYERNRGERRSHKFKF-PNV 1840
Qy 1550 -----SHRLLDGMDKREPDDPSPYLLAIWTPGTANSAQPPQCKGKASKGCMCFDETC 1604
Db 1841 TRGRGSKRVGKTSKKEPRASKAMLVTLRASQEDDDADD-----FESDYE 1886
Qy 1605 ECNSLRANQTVRGTLIPCTAMRGSPFLNGTYFQVNLFPADHESLKPIDV 1658
Db 1887 E-ESYHLAPEVANKAPFVP--VGLRSPPEVSA--QIETMEELEITVNVPDV 1934
RESULT 13
Q9H197 PRELIMINARY; PRT; 2187 AA.
AC Q9H197;
DT 01-MAR-2001 (T-EMBLrel. 16, Created)
DT 01-MAR-2001 (T-EMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (T-EMBLrel. 19, Last annotation update)
DE TENR PROTEIN (FRAGMENT).
GN TENR.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21100861; PubMed=11161782;
RA Kelter A.R., Herchenbach J., Wirth B.;
RT "The transcription factor-like nuclear regulator (TENR) contains a
RT novel 55-amino-acid motif repeated nine times and maps closely to
RT SN1.";
RL Genomics 70:315-326(2000).
DR EMBL; AJ279120; CAC21448.1; JOINED.
DR EMBL; AJ279121; CAC21448.1; JOINED.
DR EMBL; AJ279122; CAC21448.1; JOINED.
DR EMBL; AJ279123; CAC21448.1; JOINED.
DR EMBL; AJ279124; CAC21448.1; JOINED.
DR EMBL; AJ279125; CAC21448.1; JOINED.
DR EMBL; AJ279126; CAC21448.1; JOINED.
DR EMBL; AJ279127; CAC21448.1; JOINED.
DR EMBL; AJ279128; CAC21448.1; JOINED.
DR EMBL; AJ279129; CAC21448.1; JOINED.
DR EMBL; AJ279130; CAC21448.1; JOINED.
DR EMBL; AJ279131; CAC21448.1; JOINED.
DR EMBL; AJ279132; CAC21448.1; JOINED.


```

DR EMBL; AJ279133; CAC21448.1; JOINED.
DR EMBL; AJ279134; CAC21448.1; JOINED.
DR EMBL; AJ279135; CAC21448.1; JOINED.
DR EMBL; AJ279136; CAC21448.1; JOINED.
DR EMBL; AJ279137; CAC21448.1; JOINED.
DR EMBL; AJ279138; CAC21448.1; JOINED.
DR EMBL; AJ279139; CAC21448.1; JOINED.
DR EMBL; AJ279140; CAC21448.1; JOINED.
DR EMBL; AJ279141; CAC21448.1; JOINED.
DR EMBL; AJ279142; CAC21448.1; JOINED.
DR EMBL; AJ279143; CAC21448.1; JOINED.
DR EMBL; AJ279144; CAC21448.1; JOINED.
DR EMBL; AJ279145; CAC21448.1; JOINED.
DR EMBL; AJ279147; CAC21448.1; JOINED.
DR EMBL; AJ279148; CAC21448.1; JOINED.
DR EMBL; AJ279149; CAC21448.1; JOINED.
DR EMBL; AJ279150; CAC21448.1; JOINED.
DR InterPro; IPR000977; DNA_ligase.
DR InterPro; IPR001005; Myb_DNA_bind.
DR Pfam; PF00249; myb_DNA-binding; 1.
DR SMART; SM00395; SANT; 1.
DR PROSITE; PS00697; DNA_LIGASE_A1; UNKNOWN_1.
DR NON_TER 2187
SQ SEQUENCE 2187 AA; 245726 MW; 2FB083F5F82AFB55 CRC64;

Query Match      2.7%; Score 241; DB 4; Length 2187;
Best Local Similarity 18.4%; Pred. No. 2.2e-06;
Matches 345; Conservative 254; Mismatches 685; Indels 594; Gaps 82;

QY 2 Q5IMSSAVNATEQNDGSRQDVFLEFDLNTPOOKPSKRRK-----PMPKVY 51
DB 430 RSOKDAQTVEEESLTLSREDAEQVALEVDLN---QKKRRKKQDGANELGVNNLENAT 485
QY 52 VEGKP-----KRPRKPAELPKVVVGKPKRKAATQ-----EKVSKETGSA 96
DB 486 VOAGPSKGEKHKKCOAIRPE-LKEGEGSKEQMLCTQNIIDGIVGFASTKEVE-KRTDPI 543
QY 97 KKKNLKESATKPKANVGMSNK---SPEVTLK-----SCRKALNFLEN----- 137
DB 544 LSLNQDQATSVATESSESTSDLSFEVIGIRALCEVNAAGSCIEERNVDLKNLSLEID 603
QY 138 -----PGDAR-----QGDSESEIVQNSSGA-----NSF 160
DB 604 QTENVKPLMRGRFRQPKPNLSRAGKKSVLQSGKTESESKNSHKSIVSEKNHVEKDKMNTL 663
QY 161 SEIRDAIGTNGSFLDSVSQIDKTNGLAMNQ-----PLEVSMGNQPKLSTGAKLARDQ 215
DB 664 DILRMETTERENPEAETVSDLGKKNLQEGSQLKALRPVQVRGLQPKPNAGKAAERKE 723
QY 216 ----OPDL-----LTRNQOCQFPVATONTQFPMENOQAWLQMKNLIGFP-----FGNQO 261
DB 724 ILLSQEIEGANVEKNENESC-----ADRTQPMHEVDQSRKDFEEDVILQPEKNDSFQNVQ 779
QY 262 P-----RMTIRNOQCPCLAMNQOPMYLIGTPRPALVSGNQOQLGGPQGNK 305
DB 780 PDEPKVLNECLSVQENNRANKLNQVPILRTFQPKPNIGRG-----TGRREISSKEEVL 834
QY 306 RPIFLNHQ--TCLPAGNQLYGSPTDHWQLVMSTGCGQGLLKLKNQOPGSLIRGQPCVPL 363
DB 835 EKILVSGEMAAALRETVRLDTSKPMVPAEINTKEMQSDL---KETGRRAISPREKILDV 891
QY 364 IDQOPATPKGFTHLNQWATSSPSGLRPHSOSQVPTTYLHVESVSRILNGTTGTCQSR 423
DB 892 IDDTMETGLKAMREI-----CLREK 914
QY 424 AP-----AYDSLOQDIHQGNKYILSHEIS-NGNGCKKALPQNSLSLPTIMAKLEEARGSKR 478
DB 915 TPEVIDATEEIDKLEEAGR---REISPOKNGPEEVKPL-GEVETDLKATGNES-SPRE 968
QY 479 QYHRAMQOTEKHIDLNLAQIAQSQDVERHNSSTCVVEYLDAAKTKKIQKVYQNLHGMPE 538
DB 915 TPEVIDATEEIDKLEEAGR---REISPOKNGPEEVKPL-GEVETDLKATGNES-SPRE 968

```

```

DB 969 KTPECTDATEIDKNL-BETGRRKISPRENG-----PEE 1001
QY 539 VIEIEDDDPTDGARKGKNTASISKGAGSKNSPVVKKTAKEKCIIVPKTPAKKGAGRKRSV 598
DB 1002 YKPVDEMETDLNATOR-----ESSPREKTP-----VIDAT----- 1032
QY 599 PPPAHASEIQLWQP-----TPPKTPLSRSKPKGKRKSQDSGK---ARGPSGELLQCDISI 651
DB 1033 -----EEDLEETEREVSPQENGLEEVKPLGEMETDLKATGRDSFPRGKTEPEVI--DAI 1084
QY 652 AEIIYRMQNLILGDKERE--QEQNAMVILYKGDAL---VPYESKKRKPAPKV-DIDDETT 705
DB 1085 EEI-----EIDLEETEREISPOENGLEEVKPLGEMOTDLKATGREISPREKTEPEVIDAT- 1138
QY 706 RIWLLMGKGEKGEKDKKKKEKWEERVFRGRADSF IARMHLVQGDGRFRSPWKS 765
DB 1139 -----EEDKDLDEE-----TGREISP----- 1155
QY 766 VVDSVIGVFLTONVSDHLSSSAFMSLAARFPKPLSSRDERNRSVVVEDPEGCILNLN 825
DB 1156 -----EENGPEEVKPVDEMETDLKTTGREGSSREKTR-----VIDAAEVETOLE 1201
QY 826 ELPSPQEKVQHPSDMEVSGVDSGSKQELRDCNSGIERENFLEKSIQNLLEEVLSSQDSF 885
DB 1202 E-----TEREISPOENGPEE-----VKPVGKMETDLKEIREISORE--- 1238
QY 886 DPAIFQSGRGVSGSCSCSDAEFTTR-----CET-----KTVSGTSQSV----- 925
DB 1239 -----KVLAFFSAIREKEIDLKTKGRDIPIMEKVSCKMAVVEEMAD 1281
QY 926 --QTSPLN-----SDEICLQGNRPHLYEGSDGVQKQETNNVAAKKPDLEKTMNKDVC 979
DB 1282 LKETCKENFRERGSEICV-----TEEKVAELQTKGTDISPREENELETSTSR----- 1330
QY 980 FQCPRNDTNWQTPSSSYEQCATROPHVLDIEDFMQOEGLCYSWMSISPRVDRVKNKVN 1039
DB 1331 -----QTDTHLMOSGNDFSAMPS-----LDIQNISSE-----VLSMHTPVEEKRSEK 1377
QY 1040 PRRF--FRGG---SVPREFTGQIIPSTPHELPGMGLSGSSSAVQEHQDDTHNQDDM 1093
DB 1378 SSHFSHFKISSOTHESDKTEVQIOGSDVPDPOFSDINLSKSL-----PQEKPLEI 1428
QY 1094 NKASHLQKTFDLNLSSEECITRQSTSKONITDGLCPDRDAED--VVDPLSNNSLQNI 1151
DB 1429 KPAPFVRSR-----KRPKPNLARAALKRETTSEKVIYEKKSETKKMETI 1474
QY 1152 LVESNSSNKEQATVAYKETNATILREMGTLADGKK-----PTSQW----- 1192
DB 1475 VMOENNEQTDTLPSQDEASLMSR-KDTLGHRRNEEAIVLPCTQTERNLSPNSCEPK 1533
QY 1193 DSLRKDVEGNEGRQERNKNMDSIDYEAIRRASISEISEAIEKERMNNMLAVRIKDFLER 1252
DB 1534 ESQSAPVKNDVSVSGTNNVNTFOE-----WKESVIQTARQVRGLQRPRIKKTQOR 1589
QY 1253 IVKDHG---GIDLEWLRSPDKAKDYLLSIRGLGLKSECVRLTLHLNLAFPVDTNVGR 1309
DB 1590 QIVDKGEAKGIKEGRTILPKDETEK-----KVLTVSN--SQIETEIE- 1630
QY 1310 TAVRMGWTPQLPESLQLHLELYPVLESIOKFLWPRCLKDQRTLHELHQLITFGVK 1369
DB 1631 -----VPSSAVPE-----HRMYENQSVQVVLNVL 1654
QY 1370 FCTKS-----RPNACPMRGEC--RHEFASAYASARLALPAPEERSLTSATIPVPESFP 1422
DB 1655 HVNKTNETIRHENKYPVFSQAQMTRRKQOKAPNLGRHAHSKEEVPVLEKVTVDQSKGKP 1714
QY 1423 PVAIPMIELPLPLEKSLASGAPSN-----RENCEPIIEEPASPGQTEITEIDEDAY 1477
DB 1715 -----EDHLLQKGSANTQLLKEKAEKLLTSLEVSARKDCVGSRESAL--AKI 1759
QY 1478 NEDPDEI--PTKLNTEQFG-----MTLRHEMERN-----MELOEGDMSKALVALHPTTTS 1526
DB 1760 DAEELEVGPSSRRVGEETVGDNSPSSVVEEYLNKLTSCPQLNETSYSK--IALDGKTTI 1817

```


[illegible]

Qy	191	NQPLEVSMGNQPKLSTGAKLADQDPDLT---	NRQOQFPVATQNTQFPWEHQQAQLQ	247
Db	648	-----EPESGSGP-----	RPVLSYFKQFQKSLPFRFQROEQOMKQO-WQO	689
Qy	248	MKNOLIGFPEGNOQPRMTIRNOQPCLAMGNOQPMYLIGTPRALYVSGNQOGLGGPGGNKRP	307	
Db	690	QOQGVLPQTVPSPQSSVTPPPP-----	HRPLTQPMQPH-----QHLASMGFDPWR	737
Qy	308	IFLNHQTCPLPAGNOLYGSPTDMHQLVMSTGGQOGLHLIKNOOPGSLIRGOQPCVPLI--D	365	
Db	738	LMM-----QSYM DPR-----	MMSG--RPAMDIPPHPGMI-----PPKPLMRD	774
Qy	366	QOPATP--KGFTHLNQMV--ATSNSSPCL-----	RPHSOQVPTYTLHVESVSRILN	413
Db	775	OMEGSPNSSSEFHIARSADHAISLSEPRMLWSDPYPHAEPOQATTPKATSEPEDV--	832	
Qy	414	GTTGTCORSRA-----	PAYDSLQDDTHOQNKVILSHEI-----	446
Db	833	-----RSEALDQEGITAAYSVEHNQLEAHPKADFIRESSEAOVQKFLRSVEDVRPH	885	
Qy	447	-----SNGNGCKKALPONSSLPPTP---	IMAKLBEAGSKRYHRAMGQTEKKHDLNLAQOIA	499
Db	886	HTDANNOASQACFEA-PDQKTLSPAQEERIISAVESQPSRKRSVSHGSHNHTQKPDQORSEPSA	944	
Qy	500	QSQDVERHNSSTCVEYLD-----	AAKTKTKQVVQENLHGMPPVEIIEDDTDCARKGN	555
Db	945	GIPKV-----TSRCIDSKPIERPEEKPKKEGIRSEGGPKPEKVVYKSKSETRWGPRESSN	1000	
Qy	556	TASISKGASKNSPVKK-----	TAEK--EKCIVPKTPAKKGRAGR	594
Db	1001	RREEVNDPRVRRGGPIKKPVLRLDMKBEORKEKEGAEKVTEKVVV--KPEKTEKKDL	1058	
Qy	595	KKSVPHPAHASETOLWQPPPKTPLSRSPKPK-----	GRKSTQDSKGARGSGELLQDS	650
Db	1059	PPPPPPPPAPLOPOSVPVPPPTOPEAEKFPSTETATLAKQPSQDTEKPLEVPSTVQVEPA	1118	
Qy	651	IAETIYRMQNLXLGDCREQEQANMVLVYKGDGALVPYBSKKRKP--KVIDDDETT---	705	
Db	1119	VKITV--NQOTMAVPVKEQKPEKVIK-----	DLVIERPRDPSRPVAKKESTLPP	1167
Qy	706	-RIWNLMLMGDEKEGEDEKDKKKKWKWEEERRVFRGRAD-----	SFIARMLHVQ	754
Db	1168	RTYW-----	KEARERDWFPDQYGRGRGEYYSRGRSYRGSYGRGRGRGR	1212
Qy	755	GDRFESPW---KGSVDSVIGVFLTONYSDDLHSSAFMSLAAREPPKLSLSSRDEERNVR	810	
Db	1213	GHTRDYPOYRDNKPRAEHIPTSGPLRORESEETRSESSDEFV-----	PKRRQORSEDTD	1268
Qy	811	SVVVE---DPEGCILNLEIPSQEKVQHPSDMEYSGVDGSGKEOLRDCNSGIERNFL	867	
Db	1269	SEIHESASDKDS--LSKGKLPKREERPENKPKVK-----	PHSSFKPONHV	1311
Qy	868	EKSTONLEEVLSQDSFOPAI-----	FQSCGR-VGSCSCSKSDAEFFPTR	912
Db	1312	RIDNRLLEKVPYRDDKAKGFLPKGEPTPRRGGTFRRGHDGPRPSRSTLRRPAYR	1371	
Qy	913	-CETKTVSGTSQSVQTGSPNLSD---ICIQGNERPHLYEGSGDVQKQF-----	957	
Db	1372	DNQWNPQSEVPKPEDGEPRRRHEQFIPTAADKPPKFERKFPDARERPRQRPTRRPQ	1431	
Qy	958	-----	TTNVAQ-----KKPDLEKTMWKKD	976
Db	1432	DKPPFRRLREREAASKSNEVAVPTNGTVNNVAQEPVNTLGDISGNTKPTDL-SNQNSSD	1490	
Qy	977	SVCFGQPRDNTNMOT--TPSSSVBOCATRQPHV---	LDTEDFGMOGEGLYGSWMSTSPR--	1030
Db	1491	QA-----NEEWETASESDFNERRERDEKKNADLNAGTVVKGVENVLPKREIARXSF	1543	
Qy	1031	-----VDRV-----	KNKNVPRFRFRQGSVPREFTGQIIP---STPHLEPGMGL	1071
Db	1544	SSORPVDNRNRNGNGPKPSGRNFGSPRNRARG---	PPSKSGKRGPPDDOPAGTTGVLDL	1600

Qy	1072	SGSSAVQEHQDDTQHNOQDENMKASHLOKTFDLNLSSEBCLTFQSSSTQNITDGCPLR	1131
Db	1601	INGSSA-----HHQGVNPGTG--QK-----NSKSTGKKREDPK-----PG	1635
Qy	1132	DRTAEDVDPPLSN---NSSLONILVESNSSKEQTAVEYKETNATILREMKGTLADGKKP	1188
Db	1636	PKPKPKKYDALSQFDLNNYASVVID---DIPEVTVIDPQSN---LNDGFTVWSKKQ	1689
Qy	1189	TSQWDSLRDKVEGNEGRQBR-----NKNMNSDIYEAIRRASISEISEAIKRGNNNM	1241
Db	1690	-----QKRLODERKKKEQVIOWNKKANE-----KGRSQTSK	1724
Qy	1242	LAVRIKDFLERIVKHGGIDLEWLRESPPDKAKDYLLSIRGLGLKSECVRLTLTLHNLAF	1301
Db	1725	L-----PFRFAKKQATGQ-----QAQSSASVPPPLASAPL	1754
Qy	1302	PVDNTGRIAVRMGWVP---LOPLPESLOLHLELPVLESIOKFLWPRLCKLDQRTLVEL	1359
Db	1755	PPSTSASVPASTSAPLPAFLTPVPASTSA-----PVPAS-----	1788
Qy	1360	HYOLITFGVKFTKSRPNCAQPMGECRHPASAYASARLALPAPERSTSATIPVPE	1419
Db	1789	-----TLAPVLASTAP-VPASPL-----APVSASASVSVPASTSAAAITSSAPA	1835
Qy	1420	SFP-----PVAIDPMI---ELPLPLEKSLAS-CAPSNNRENCPIITEEPASPCQEC	1464
Db	1836	SAPATPILASVSTPASVTIILASAPI-LASALASTSAPT-----APASSSNAP-----	1886
Qy	1465	TEITESDIEDAYNEDPDEIPTIKLIEQFGMTLREHMERMNELOEGDMSKALVALHPTT	1524
Db	1887	-VITAPTI-----PASAPT-----ASVPLAPAS	1908
Qy	1525	TSIPTPKLKNISRLRTEHOVYELPDSHRLLDGMDKREDDPDPSPYLLAIWTPGETANSAP	1584
Db	1909	ASAPAPAPTFS-----APNPAPPAP-----AQTAQAQTHKPVQN	1942
Qy	1585	PEQKCGKGASGMCFDETCSECN-----SLREANSQTVRGFTLLPICRTAMRGSPLNGTY	1639
Db	1943	PLQ-----TTSQSKQPPPSIRLPSAQTTNGT-----	1969
Qy	1640	FQWNELPADHESLKPDIIVPRDW-----IWD	1665
Db	1970	-----DYVASGKSIOTPQSHGTLTAELWD	1993

Search completed: September 27, 2002, 13:13:22
Job time: 1115 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 27, 2002, 10:04:11 ; Search time 238.43 Seconds
(without alignments)
805.463 Million cell updates/sec

Title: US-09-840-743-2

Perfect score: 9089

Sequence: 1 MQSIMSSAVNATEQND.....PRPLMARLHPFASKLNKNT 1729

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_032802.*

```

1: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	246.5	2.7	3256	21 AAY50976	Human cell cycle p
2	244.5	2.7	2724	22 ABG20119	Novel human diagno
3	237	2.6	2819	22 AAB35408	Human 07CG27 gene
4	232	2.6	2290	22 ABB60066	Drosophila melanog
5	230.5	2.5	2063	22 AAM78710	Human protein seq
6	230.5	2.5	2063	22 AAM40064	Human polypeptide
7	230	2.5	3263	22 ABB67210	Drosophila melanog
8	225.5	2.5	2063	22 ABG17147	Novel human diagno
9	225.5	2.5	2519	22 ABG16636	Novel human diagno
10	225.5	2.5	2759	22 ABG17148	Novel human diagno
11	222.5	2.4	5447	22 AAU14697	Novel bone marrow

12	222	2.4	3257	22 ABB67502	Drosophila melanog
13	221.5	2.4	5533	22 ABB65772	Drosophila melanog
14	221.5	2.4	5560	22 ABB71160	Drosophila melanog
15	216	2.4	5373	22 AAU14603	Novel bone marrow
16	214.5	2.4	2362	22 AAB58396	Drosophila melanog
17	213	2.3	2263	22 AAM79000	Human protein seq
18	211.5	2.3	2439	22 ABB68896	Drosophila melanog
19	210.5	2.3	1552	22 ABB71764	Drosophila melanog
20	210.5	2.3	1805	13 AAR27204	Rat nestin. Rattu
21	210.5	2.3	1805	15 AAR60126	Rat nestin. Rattu
22	210	2.3	1852	22 ABB67170	Drosophila melanog
23	210	2.3	2016	22 ABB71487	Drosophila melanog
24	208.5	2.3	2432	21 AAY85565	Human homologue of
25	207	2.3	1637	22 AAM00916	Human bone marrow
26	207	2.3	2951	22 ABB60291	Drosophila melanog
27	205.5	2.3	3111	22 ABB60327	Drosophila melanog
28	205	2.3	1982	22 ABG16404	Novel human diagno
29	205	2.3	1982	22 ABG19656	Novel human diagno
30	205	2.3	6815	22 ABB68811	Drosophila melanog
31	204.5	2.2	2194	22 AAM40114	Human polypeptide
32	201	2.2	1703	22 ABB66223	Drosophila melanog
33	201	2.2	3238	22 ABB71715	Drosophila melanog
34	201	2.2	5024	22 AAG82935	S. epidermidis ope
35	200.5	2.2	2703	22 ABB60074	Drosophila melanog
36	198.5	2.2	1920	22 ABB64441	Drosophila melanog
37	198.5	2.2	1976	22 ABB69419	Drosophila melanog
38	198	2.2	1984	22 ABB61060	Drosophila melanog
39	197.5	2.2	2400	22 ABG20278	Novel human diagno
40	197.5	2.2	3201	22 ABB62899	Drosophila melanog
41	196.5	2.2	1795	22 ABG21018	Novel human diagno
42	196	2.2	2189	22 AAM79017	Human protein seq
43	196	2.2	2285	22 ABB63057	Drosophila melanog
44	195.5	2.2	2161	22 AAM78959	Human protein seq
45	195.5	2.2	2523	22 AAU03503	Human protein kina

ALIGNMENTS

```

RESULT 1
AAY50976
ID AAY50976 standard; Protein; 3256 AA.
XX AC AAY50976;
XX XX
DF 10-MAR-2000 (first entry)
XX
DE Human cell cycle protein Ki-67.
XX
KW Cell cycle protein; Ki-67; therapy; cell proliferation; allergy;
KW tumor treatment; autoimmune disease; scar formation; inflammation;
KW rheumatic disease; transplantation.
XX
OS Homo sapiens.
XX
PN DE19822954-A1.
XX
PD 25-NOV-1999.
XX
PF 22-MAY-1998; 98DE-1022954.
XX
PR 22-MAY-1998; 98DE-1022954.
XX
PA (BORS-) FORSCHUNGSZENTRUM BORSTEL ZENT MEDIZIN.
XX
PI Flad H, Gerdes J, Boehle A, Deinert I;
XX WPI; 2000-039964/04.
DR N-PSDB; AA43872.
XX
PT Ki-67 gene antisense oligonucleotide -
XX Disclosure; Page 19-29; 36pp; German.
PS

```

xx This invention describes a novel oligoribo- or oligodeoxyribonucleotide, CC characterized in that, it hybridizes to mRNA that encodes protein Ki-67 CC at a physiologically acceptable salt concentration. The oligoribo- or CC oligodeoxyribonucleotide which is complementary to Ki-67, a protein CC active at all stages of the cell cycle except G₀, is useful for therapy CC of illnesses with increased cell proliferation and particularly for CC treatment of tumors, autoimmune diseases, scar formation, inflammation, CC allergy, rheumatic diseases and defence against transplantation. This CC sequence represents the human cell cycle protein Ki-67 which is described CC in the method of the invention.

xx
SQ Sequence 3256 AA;

Query Match 2.7%; Score 246.5; DB 21; Length 3256;
Best Local Similarity 18.8%; Pred. No. 6e-09;
Matches 362; Conservative 266; Mismatches 768; Indels 527; Gaps 94

QY 22 SRODVLFEFLUNKTPQOKPKRKRKMPKVVVBGK----PKRKP- -KPAELPKVVGEGK 75
DB 1496 Sqpdpvdtstsskpskrsr-----kvdveeffalrkrtpsgakamhtpkpavgek 1549
QY 76 KRKPRAATQEKVYSKETGSAAKKNI-----KESA-----TKKANVGDMSNK 118
DB 1550 niyafmtpvgkidltenitgskrrlqtptkekaqadlagfkelfqtrghteesmtndk 1609
QY 119 SPEVTLKSCRKALNFDLEPNPGDARGDSESETVQNSGANSFSEIRDAIG--GTNG--- 172
DB 1610 takvackssqpdid---knpasskrr-----lkslkgkvgeellavglkltsgett 1660
QY 173 -----SFLDSVSI-DKTNGLGAMNOPLEVSMGNQ---PKLSTGAKLARDQ 215
DB 1661 htchteptgdgksmkafmespkildsaasltgskrqrlrtpkgsvepedlagfielf--q 1718
QY 216 QPDLLLTRNOCCQFPVATQNTQFPWENQQ---AWLQMKNLGIFGFEQNO-QPRMTIR-- 267
DB 1719 tp-----shcksmntektkvsyrasqpdilvdtptsskqpkqsrirkad 1763
QY 268 NOQPCLAGNQOPM--YLIGTPRALVSGNQQLGGPQGNKRPIFLNHOITCLPAGNQLYGS 325
DB 1764 teeflafrktpsgakamhtpkpav-----geekdi-----ntflgt 1801
QY 326 PTDHQLVMTSGGQGHLLTKNOQPSL--IRG-----QOPCV--PLIDQOPATPKGTH 376
DB 1802 pvqkldgpgnlpgsnrrlqtrkekaqaleeltgfrlftqptcdhptadek-----tt 1854
QY 377 LNQWVATSMSPGLRPHSQSVPTTYLHVBSVS-----RIILNGTGTGQSRAPAYDSL 430
DB 1855 kllicksqsdapdtntkqrkrsllkadveeefflafrkltsgakamhtpkpavge- 1913
QY 431 QODIHQ--GNKYILHSIFISNGCKKALPONSSLPTPIMAKLEEARGSKRQY----HRAM 484
DB 1914 ekdintfvtgtpvekldilnlpqskrr-pqt---pkekakaledlagfkelftqpghtee 1969
QY 485 GOTEKHDNLNLAQQTAAQSQDVERHNSSTCYEYLDAAKTKTQKVVOENLHGMPEVETED 544
DB 1970 smtdkdtvecksqpdqpvktpstss-----kqrlkislgkv-----gvkeevlpv-- 2015
QY 545 DPTDARGKNTASTISGA-----SKGNSSPVRKTAKEKCIY-----PKTPA 587
DB 2016 -----gkltqsgkttqthretagdgksikafkesakqmidpanygtgmerwprtpk 2067
QY 588 KKGK-----AGKKSVVPPAHASE-----IQLWQPTP-----PKTPLSR 621
DB 2068 eeagsledlagfkelftpdhteeesttdkttkackspppsmdtptsttrrpktpigk 2127
QY 622 -----SKPKG---KGRKSIQDSGK-----ARGPSGE 644
DB 2128 rdiveelsalkqltqthtdkvpqdedgkinvretakqldpaasvtgskqrptpkqk 2187
QY 645 LLCODSIAEIIYRMQNLVIGBKEREQBNAMVLYKGD-----GALVPYESKKRPRPKVD 699

Best Local Similarity 19.18; Pred. No. 4.6e-08;
Matches 356; Conservative 208; Mismatches 668; Indels 636; Gaps 82;

```

Qy 40 SKRRKFKPKVVEGCKRKRKPA---ELPKVV--VEGPKRKRK-KAATQEKVKRSKE 92
Db 8 nkrrk-----etelpppppvpagsslapsvaaiesppmpkrktkvaqrkfaqq 62
Qy 93 TGSARKKNLKESAT-----KKPANVGDMSNKGSP-EVTLKSCRKALNF----- 133
Db 63 sapssssaiysaaaaagsgpstgsadnqpdiellpnkcpkqdfllfclfrgtra 122
Qy 134 -----DLNPGDARGDSESEIVONSSGANSFSEIRDAIGTNGSFLDSVQIDKTNGLG 188
Db 123 lpahldfInqgskdseasstgannnnarKpnnsraslaguk-----kkrgrp 171
Qy 189 AMNQPLEYS---MGNOPDKLSTGAKLARDQOPDLLTRNQOCQFPVATQNTQFPMENQQA 244
Db 172 akgatckvddvwapldptcktpaekpasd----- 201
Qy 245 WLOMKNOLIGFPFGNQOPRMTIRNQOPLAMGNOOPMYLIGTPR---PALVSGNQQLGGP 301
Db 202 ---knnli---aaeapkiprntptvlp-----gavrkraevvldgnrr--ga 241
Qy 302 QGNKRPIFLNQTCLPAGNQLYGSEPTDMLQVLMSTGGQOGLLLKNNQOPGSLIRGQPCV 361
Db 242 rg--rgagishdkttpa-----lkdnkrrsnrastkesk 275
Qy 362 PLI-----DOOPATPKGTHLNMVATSMSSPGLRPHSOSOVPTTYLHVESVSR--I 411
Db 276 pviddnnedldegeaddedefsahdatrsngeevkadsvpkaepckrprsaakav 335
Qy 412 LNTGTG--TCORSRAPAYDSLOQDIHQGNKYILSHIEISNGCKKALPONSSLPTPIMAKL 470
Db 336 epaegqvtlkdapkvrvqr---espifipsessgrmrtraffepkvvpvh----- 387
Qy 471 BEARGS-----KQYHRAMGQTEKHDLNLAQOIAQSODVERHNSSTCYEYLDAAKT 522
Db 388 keepahdpdkdepkkqpetsktek--egetpkqikeak--ektnssi-----yakea 438
Qy 523 KIQKVQVQNLHG---MPEVETIEDPDGARKGKNTASISGKASKGNSPVKKTAEK 579
Db 439 kqqlengdsllhgavakgkksinvfssdsdeqplakskllkga--kg--aspakeskvee- 495
Qy 580 CIVPTPAKGRAGRKKSVPPPAHASEIQLWQTPPKTPLRSKPKGKRKSIODSGAR 639
Db 496 ---pbggkrgkagkkaedapadeelI-----appsskkvatnarg--srksak--skas 544
Qy 640 GPSGELLQDS-----IAEIIYR---MQNLXLGD-----KEREQ 671
Db 545 edsaeleprsqrparktkaaalymgllghklqladdeedlsmssfpdipnvkemekm 604
Qy 672 QNAMLVYKGDGALVPYES-----KKRKRPRK---VDIDDETT----- 705
Db 605 eneiKnaagklnvpeastvlsagstkprkprkdsptvtaetrnespsppeaeek 664
Qy 706 -----RIWNLKMGKDEGEDEKKKKEKKEWEEERRVPRG 741
Db 665 pppparrrpqbqkvnapghkpriegmlvkkgslakmsseqgakpkak---sehsksvdsd 722
Qy 742 RADSPIARMHLVQGDPR---FS-----PWKGSVVD 768
Db 723 eeedflineinetkrlekfsdsdddeplaikvpavavkekellppkmaspppklvvpv 782
Qy 769 SVIGV-----PLTONVSDHLSS--SAFMSLAARFPKP 798
Db 783 spisvptpavsvapslavpsilptstgttqltmpltkatatslgsmmpfisiqasvpl 842
Qy 799 LSSSREDERNVRSVVEDEPGCILNL-----NEIPSWQEKVQH----- 836
Db 843 ktmekkp-----pyptskilnvpatalpgsasgavsfqktsylvpgasphyhsy 893
Qy 837 -----PSDMEVSGVDSKQELRDCS-----NSGIERE-----NFKESIQ 872

```

RESULT 5

AAW78710

ID AAW78710 standard; Protein; 2063 AA.

XX

AAW78710;

XX

DT 06-NOV-2001 (first entry)

XX

DE Human protein SEQ ID NO 1372.

XX

Human; cytokine; cell proliferation; cell differentiation; gene therapy;
vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
tissue growth factor; immunomodulatory; cancer; leukaemia;
nervous system disorder; arthritis; inflammation.

```

Db 894 vrpptpthqfggagpagsktpsgssplkyqtpttpttpppieayqcpkinpnfl---tp 950
Qy 873 NLEBEVL--SSQDFDPAIFQSCGRVGSCKSDAEFPPTTRCETKTVSGTSQS----- 924
Db 951 kyersplprsstsfp-----spspvkfqptsqgttattvgtppssatkspl 995
Qy 925 -----VOTGSPNLSDICLOGNERPHLYEGSGDVQKQETTNVAQKPDLEKT 971
Db 996 apppppsvqvigtsslnlmslaatvnaap-----mpvastaaaaahgttvatt 1047
Qy 972 MNWKDSVFCGQPRNDTNWQITPSSSYEQCATRQPHVLIEDFGMGEGGLGYSWMSISPRV 1031
Db 1048 v-----tatstppahnsaat-----ggag---ansdpik 1076
Qy 1032 DRVKNKNVRRFFRGGSVPREFTGIIIPSTPHEPLPGMGLSGSSSAVOEHQDDTQHNOOD 1091
Db 1077 dei-----gsilaqat--lmpskeesgkifgiavslaqsagpdnktct--- 1118
Qy 1092 EMNKASHLQKTF-----DLNSEECLTROSSTKONITDGCCLPRDRTAEVVD 1140
Db 1119 -lgcgsihkpvlgpvvptegyfgdqlsskerrkakvnmtheqikwliecssnpdeiqd 1177
Qy 1141 PLSNNSSLONILVESNNSKQETAWEYKETNATILREMKGTGLADGKPTQSDSLRKDVE 1200
Db 1178 dladd--fdslrppqstppptr--dkelsasfsssskntrgdlgkesawsakgpsik 1234
Qy 1201 G-----NEGRQERNKNMDSIDYEAIRRASISISEAIRKERNMNLAVRIKDFL 1250
Db 1235 atpvlvtpnrrkelheqadskdcaldyd--ksstpvnlqtkits---neslav----- 1283
Qy 1251 ERIYKHGGIDLEWLRESPPDKAKDYLLSIRGLG-----LKSVECVRLLLTLHNLAFPV- 1303
Db 1284 ekkvndrk-----kesakaaasakaatqptrstaatpbtptissttpaslt--pskssptp 1336
Qy 1304 -----DTNVGRIAVRMGWVPLQLPESLQLHLLLELYPVLESIQKFLWPRLCK 1350
Db 1337 ppavkqkaekkrnatagggatalaspaaaptpanpk-----rlpv 1377
Qy 1351 LDQRTLYELHVLQITFGVFCTKSRPNCNACPMRGECRHFASAYASAR----- 1398
Db 1378 ynknaqaqqq-----aetkpsan--ppsgagtkresvayafgkddesgsksnrr 1426
Qy 1399 -----LALPAP-----EERSLTSATIPVPESPPVPAIPMIELPL-PLEKSLASG---- 1442
Db 1427 rtdpsvpapalvpnlalsers-----ptkrraaaaataatvqltslptenckiegkpsk 1480
Qy 1443 APSNRENCPIIEEPASPGQECTETTESDIEDAYV----- 1477
Db 1481 aptgrgakkqgqapapappvpeasgdsdaegatfiplqgavvgsgdggigqvavklg 1540
Qy 1478 NEDPDEIPTIKLNIEQFGMTLREHMERME-----LOEGDMSKALVALHPTTTTPTPKL 1532
Db 1541 regpdg--pnqkv--vmqatlvtkagmdtnskplpeslnelvktilhaasndaaatttsl 1598
Qy 1533 KNISRLRT 1540
Db 1599 kslpkast 1606

```


QY 1237 GNMNLAVRIKDFLERIVKHGGIDLEWLRSPDPKADKDYLLSIRGLGLSKSVECVRLTL 1296
 Db 1499 gapv-----tikppgltdle---vttp-----vvsgeclkkasvip--tl 1534
 QY 1297 HNLAFPVDVTNVRTAVRMGVPLQPLPESLQLHLELYPVLESIOKFLWRLCKLDORTL 1356
 Db 1335 qdlsskepsns-----lnlphsnelcsslvhpelsevssnvapsipvmrsv 1583
 QY 1357 YE-----LHYQLITFGVCTKSRPNACPMRGECHFAAYASARLALPAPPEERSLT 1410
 Db 1584 ssssislplpnqit---vftv-snp-----ittsantaaalpthlqsal 1625
 QY 1411 SATIPVPESPPVAIPMIELPLEKSLASGAPSNRENCEPIIEEPASQCEITEITES 1470
 Db 1626 stvtmnpag-----skvmvseggaagsnarpqfittpvf-----inss 1664
 QY 1471 DIEDAYNEDDEPTIKLNT-----QFGMTLRHMERNMELQEGMSKALVALPHTT 1524
 Db 1665 sliqmkgsqst.ipaaplttnslmpsvavvgplhlpqnik-----fsapvpphals 1719
 QY 1525 TSITPKLKN-----ISRLRTEHQVYELP-----DSHRLLDGMDKREPDPSPYL--- 1569
 Db 1720 ss-papniqtrplvlssratpqlpspctsspvvpshpvpqvkelnpdeaspqvnts 1778
 QY 1570 -----LAIWTPGETANSAQPPQKCG-GRAS--GKMCFTETCSE-CNS 1608
 Db 1779 adqntlpssqgtmvspltnspgssgnrrrsvsssgkgkvdkgigqlltkackkvts 1838
 QY 1609 LREANSQ 1615
 Db 1839 lekgeeq 1845

RESULT 6
 ID AAM40064 standard; Protein; 2063 AA.
 XX AC AAM40064;
 XX DT 22-OCT-2001 (first entry)
 XX DE Human polypeptide SEQ ID NO 3209.
 XX KW Human; neurotropic; immunosuppressant; cytostatic; gene therapy; cancer;
 KW peripheral nervous system; neuropathy; central nervous system; CNS;
 KW Alzheimer's; parkinson's disease; Huntington's disease; haemostatic;
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
 KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
 KW leukaemia.
 XX OS Homo sapiens.
 XX PN WO200153312-A1.
 XX PD 26-JUL-2001.
 XX PF 26-DEC-2000; 2000WO-US34263.
 XX PR 21-JAN-2000; 2000US-0488725.
 XX PR 25-APR-2000; 2000US-0552317.
 XX PR 09-JUL-2000; 2000US-0598042.
 XX PR 19-JUL-2000; 2000US-0620312.
 XX PR 03-AUG-2000; 2000US-0653450.
 XX PR 14-SEP-2000; 2000US-0662191.
 XX PR 19-OCT-2000; 2000US-0693036.
 XX PR 29-NOV-2000; 2000US-0727344.
 XX PA (HYSE-) HYSEQ INC.
 XX PI Tang YF, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
 PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
 PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
 XX

DR WPI; 2001-442253/47.
 DR N-PSDB; AAI59220.
 XX Novel nucleic acids and polypeptides, useful for treating disorders
 PT such as central nervous system injuries -
 XX Example 5; SEQ ID NO 3209; 10078pp; English.
 XX The invention relates to human nucleic acids (AAI57798-AAI61369) and
 CC the encoded polypeptides (AAM38642-AAAM42213) with neurotropic,
 CC immunosuppressant and cytostatic activity. The polynucleotides are useful
 CC in gene therapy. A composition containing a polypeptide or polynucleotide
 CC system, such as peripheral nervous injuries, peripheral neuropathy and
 CC localised neuropathies and central nervous system diseases, such as
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
 CC utilisation of the activities such as: Immune system suppression,
 CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
 CC assays for receptor activity, arthritis and inflammation, leukaemia and
 CC C.N.S disorders.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification.
 XX Sequence 2063 AA;
 SQ

Query Match 2.5%; Score 230.5; DB 22; Length 2063;
 Best Local Similarity 18.0%; Pred. No. 5.1e-08;
 Matches 326; Conservative 225; Mismatches 645; Indels 611; Gaps 80;

QY 52 VEGPKPKRP-----RKPAELPKVVEGPKPKPKRAATQEKV----- 88
 Db 407 lggpsrvptlqqphltknspasspsfqgspassstvnqtqgmgprppqnnlpqg 466
 QY 89 -----KSKETG-SAKKKNLKESATKKPANV 112
 Db 467 fqqpvssgrnrpmvqggnvppnfmvmgqgppnqppqslhpglgnpkrllppgfsagqan 526
 QY 113 GDMSNKSPVTLKSCRKA-----LNFLENPGDARQDSESEIVQNSGANSFSE----- 162
 Db 527 nfmaggvptstatlpgnsgapqlqanqnvqhagggagppqng-mqvshgppnmqpslm 585
 QY 163 -----IRDAIGTGNSFLDSVSQDKTNGLGAMNQ-PLEVSMGNQPKLSTGAKLARQQ 216
 Db 586 gihgmmnnqgatsg-----vpqvnlsnmqgqgppsqimgmhqqlvpsgggmvgqg- 639
 QY 217 PDLLTRNOCCQFPVATQNTQFPMENQQALQMKNLIGFP-----FGNQQPRMTIRNOQPC 272
 Db 640 -----gtlnpqpmllsraqlnpqgqgmvmvppsqnlgsppqgmt----- 678
 QY 273 LAMGNQOPMYLIGTRPRALVSGNQOLGGP-----QGNKRPIFLNH 312
 Db 679 -----ppkqmlsqggpqqmaphuqmgpgqgvlqqnqmiegimtnqmgknkqfntqn 732
 QY 313 QT-CLPAGNQLYGSPT-DMHQLVMSTGCGQHQHLLIKNQOP-----GSLIRQQ 358
 Db 733 qsnvmppgaqimrgptnmgqnmvqftqgmsgqmlpqgppvnnspssqvmgiqgvlrppg 792
 QY 359 PCVPLIDQOPATPKGFTHLNQMVATSMSSPGLRPHISOQVPTTYLHVESVSRILNGTTGT 418
 Db 793 ps-phmaqqhgdpa---ttanndvslsqmmpdvsiaqtdtmvp----- 830
 QY 419 CORSRAPAYDSLQODIHQGNKYILSHEIS-----NGN-----GCKKALPQNS--SLP 463
 Db 831 -----phvqamqngusasn-hfsgghmsfnafsgapngngmqscgpnpgfavnkdvlt 883
 QY 464 TPIMAKLEEARGSKQYHRAMQOTEKHDNLAAQQAQTAQSDQVDERHNSSTCVELYDAAKTK 523
 Db 884 splllvllqsdlsaghf-----gvnnkquntnanpkkkkppr 921
 QY 524 IQKVVQENLH---GMPPEVIEDDPTDGARKGKNTASISKGASKGNSSPYKKTAEKEC 580

Db 148 atpiekkpapakleaakvvlksikterdggimslatleqiagkteakeaipwitmrekl 207
Qy 129 KAL-----NFDLE-----NPGDARQGDSESEIVQNSGANSFSEIRDAIGGTNGSFIL 176
Db 208 kavesvqqnlkfddevylqpleg-qieteqqlpqaaq-----ve 247
Qy 177 SVSQIDKTNGLGAMNOPLEVSMGNQDPKXISTGAKIARDQOQDPLLRNQ-OCOFFPVATQNT 235
Db 248 qvqrkteigrilksmesveImemtgdiklitqgnakdliipwkemrqqlksvgrvtkqid 307
Qy 236 QPFMEQNQAWLQMKNLQIGFPNQNQPRWTI-----RNOQPCLAMGNOQ 279
Db 308 kfkieveirhlaqagaiteeyqtgaetevmideskgsiskvlrrdeq-lyededs 365
Qy 280 PMT---LIGT-----PRPALVSGNQQLGGPQGNKRPIFLNHQTCCLPAGNOL 322
Db 366 niyqkfittedvnlmhvsaerekleaqrliREQAVNWRGQGRP-----q 412
Qy 323 YGSP-TDMHQLVMSTGGQHGLIKKQOQGSILRGQ-QPCVPLIDQOPATPKGFTH---L 377
Db 413 --qpltsvedvisqtsrqkly---ggqsflieeaqrqgvqvedsgmmsleeyehqkii 467
Qy 378 NOMV---ATSMSPGLRPHSOSOV-PTTYLHVE-----SVSRILNGTGTGCRSR----A 424
Db 468 nqrtaqeafrwqp-repqkfiqvdstllhlqerhdtceqgillqqpvmwdrgrkkp dq 526
Qy 425 PAYDSLQDD-----IHQGNKYILSHE-----TSNGNGCKKALPQNSSL- 462
Db 527 pqyvqqeqvkeefvekpkyteemhdleptieqppvpvmwergkkkppqpkte 586
Qy 463 -----PPTIMAKLEAERGSRQVHRAMGQTEKHDNLNAQIAQ-SODVERHNSSTCV 513
Db 587 eahdelveptpv---qqpevpvmwergkkkvaqgetvlsqevvqtsvveq----- 635
Qy 514 EYLDAAKTKIOKVQENLHGMPEVTEIEDDPTDGARKGNTASTSGASKGNSSPVK 573
Db 636 qiweetkktavrrvipprepqkvqvtlktprprprkeavkaeeiqlklr-strvpvq 694
Qy 574 TAEREKCI-----VPKTPAKKGRAGRKSVPPPAHASEIOLMOTPPKT----- 617
Db 695 pveaeqkayeeatdelteepipqpvmwergkkkppkqeevteI-----pktleia 747
Qy 618 --PLSRSKPK-----GGRKSIQ--DSGKARGPSGELLQCQSIATFIYRMQNL 662
Db 748 vdtleevpkptepqpqvliwargqkpkpdkqelpklsleavdtieedlik----- 802
Qy 663 LGDEREQEQNAMVLYKGDGALVPYSGKKRPRP-----KVIDDETTT----- 706
Db 803 -----pvqpeppvI-----werkkkppqpdvleekldvaptkyekavdvilpde 848
Qy 707 -----IWNLLMGKGDEKGEDEKD-----KKKEKWWEERRVRFRGR-- 742
Db 849 pkveekpepvliwqrgkkkikpsepteevhpdvdaqietvvkdedemiveekrriktkrp 908
Qy 743 --ADSFARMHLVGDGRFPWKSVDVSVICVLTQNVSDHLSSAFMSLAARFPKLS 800
Db 909 kstkevtelfeedpeeeispeeevpqkevi-----eeieeiveekrriktktkk--pklt 961
Qy 801 SSRDERNVRSVVYDDPGC TLNLNETPSMOKVQHPHSDMEVSGVDSGSKPQLDCSNSG 860
Db 962 qqvteeetpheeilikeseevvqeeeliveekkvk-----kvkpkptvaeqqlk----- 1010
Qy 861 IERFNFLKSIQNLFEELSSQDSFDPALFQSGCRVSGSCSKSDAEPTTTRCTKTVSG 920
Db 1011 -----eeelpt-eetveeetaedqqlvve-----eskkvkk 1041
Qy 921 TSQSVQTCSPNLSBICLOGNERP-----HLYEGSGDV-----QK 955
Db 1042 vkk--ptgtvktvdveelpgeevpveevpvedvapeeeelieeqeeivdqddeiqgk 1099
Qy 956 QETTNVAAOKPDLKTMWKSVCVCPQRNDTNWQTTSSSYEQCATRQPHVLDIEDFGM 1015
Db 1100 rkvvkakkpkktiekt---eieeedqpeee-----vleeeiige 1136

Qy 1016 QGEGLGYSMISPRVDRVKNKNNVPRFRFROGGSVPREFTGQIIPSTPHELPCMGLSGSS 1075
Db 1137 qee-----lterqrvksikkpkkv----- 1156
Qy 1076 SAVQHQDDTOHNOQDEMKNKASHLOKTFLD-----LLNSSECLTRQSTKON 1123
Db 1157 -vtektvdteqpeesqaeevketvteepkpkpapeakveqvekslkpaprkg 1215
Qy 1124 ITDGCLPRDRTAEVDVPLSNNSLQNLIVESNSSNKE--QTAVEYKE---TNAYILRE 1177
Db 1216 l-----lpekeqveevl-----lkpvkkiavaseaeqpetefevkefaifttedildv 1266
Qy 1178 MGTTLADGKKPTSSOWDSLKRDVEGREGROBRNNKNNMDSIDYEAIRRASISE--A 1232
Db 1267 tkkrvkk-kkptk-----vaaaesteepaeet-eefeeatqpeevqpveeipeeq 1317
Qy 1233 IKERGMNNLAVRIKDFLERIVKHGGIDLEWLRESPPPKAKDYLLSIRGLGLKSECVR 1292
Db 1318 vkevaderktapkpkrkeeliekveevalk--rvtrpkkeipqeatieevrikptg--- 1372
Qy 1293 LLTLHLNLAFPVDNNGRIAVRMGWVPLQ-----p 1321
Db 1373 -----rtsikpeevkleevdlqhvkkedeivgeekrtrkvkpkphedipeip 1421
Qy 1322 LPESLQHLLE-----LYPVLESIOKFLWPRCLKDORTLY 1357
Db 1422 daeptqlaeaeihielekqkpeedqpvpwkrgeekqpveevleekkw----- 1470
Qy 1358 ELHYQLITFGKVCTKSRPNCNACPMRGECRHPASAYASARLALPAP-----ERSLTSAT 1413
Db 1471 -----sgrrrIpeqgqeevqI-----pipskpieeqgkpeka 1504
Qy 1414 IPVp---PESFPVPAIPMIEL-PLPL-----EKSLASGAPSNRENCEP 1452
Db 1505 ipdpqlvpeekpeseeeeeIeplkIpedkkpkepkakkekxxxpkkkatcpsvdesee 1564
Qy 1453 IIE---EPASPGQECTEITESDIEDAYNED--PDE--IPTIKLINIEQFGMTLREHMERN 1505
Db 1565 vaepfdepaeeadeveempvddvkvvavsedvIpeeevvpt-----eetpeakqkakkrt 1620
Qy 1506 MELQEGDM 1513
Db 1621 krIkeasv 1628
RESULT 8
ABG17147
ID ABG17147 standard; Protein: 2063 AA.
XX
AC ABG17147;
XX
DT 18-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #17138.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX (HYSE-) HYSEQ INC.
XX Drmanac RT, Liu C, Tang YT;
PI

Db 1626 stvtvmpnag-----skvmvsegqaaqsnarpqfitpvf-----inss 1664

Qy 1471 DIEDAYNEDPDEIPTIKLNE-----QFCMTLREHMERMEIQEGDMSKALVALHPTT 1524

Db 1665 siiqvmkysqstipaapltnsglmpsvavvgpllipqnik-----fssapvppnals 1719

Qy 1525 TSITPPKLKN-----ISRLRTEHQVYELP-----DSHRLLDGMDKREDDPSPYL--- 1569

Db 1720 ss-papniqtgrpvlssratpqlpspctssvpvshppvqqvkelnpdeaspqvnts 1778

Qy 1570 -----LAIWTPGETANSAQPPQKCG-GKAS--GKMCDFETCSE-CNS 1608

Db 1779 adqntlpsqsttmvplntnspgssgnrrspvsskgkgkvdkigqlltlackkvtgs 1838

Qy 1609 LREANSQ 1615

Db 1839 lekgeeq 1845

RESULT 9

ID ABG16636

XX ABG16636 standard; Protein; 2519 AA.

AC ABG16636;

XX 18-FEB-2002 (first entry)

DT Novel human diagnostic protein #16627.

DE Human; chromosome mapping; gene mapping; gene therapy; forensic;

KW food supplement; medical imaging; diagnostic; genetic disorder.

KW Homo sapiens.

OS WO200175067-A2.

XX 11-OCT-2001.

PD 30-MAR-2001; 2001WO-US08631.

PF 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.

XX (HYSE-) HYSEQ INC.

XX Drmanac RT, Liu C, Tang YT;

PI WPI; 2001-639362/73.

XX N-P5DB; AAS80823.

DR New isolated polynucleotide and encoded polypeptides, useful in

PT diagnostics, forensics, gene mapping, identification of mutations

PT responsible for genetic disorders or other traits and to assess

PT biodiversity

XX Claim 20; SEQ ID No 46995; 103pp; English.

PS The invention relates to isolated polynucleotide (I) and

CC polypeptide (II) sequences. (I) is useful as hybridisation probes,

CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome

CC and gene mapping, and in recombinant production of (II). The

CC polynucleotides are also used in diagnostics as expressed sequence tags

CC for identifying expressed genes. (I) is useful in gene therapy techniques

CC to restore normal activity of (II) or to treat disease states involving

CC (II). (ii) is useful for generating antibodies against it, detecting or

CC quantitating a polypeptide in tissue, as molecular weight markers and as

CC a food supplement. (II) and its binding partners are useful in medical

CC imaging of sites expressing (II). (I) and (II) are useful for treating

CC disorders involving aberrant protein expression or biological activity.

CC The polypeptide and polynucleotide sequences have applications in

CC diagnostics, forensics, gene mapping, identification of mutations

CC responsible for genetic disorders or other traits to assess biodiversity

CC and to produce other types of data and products dependent on DNA and

CC amino acid sequences. ABG00010-ABG30377 represent novel human

CC diagnostic amino acid sequences of the invention.

CC Note: The sequence data for this patent did not appear in the printed

CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 2519 AA;

SQ

Query Match 2.5%; Score 225.5; DB 22; Length 2519;

Best Local Similarity 18.1%; Pred. No. 1.8e-07;

Matches 372; Conservative 287; Mismatches 787; Indels 611; Gaps 90;

Qy 25 DVLEFDLNKTPQOKPSKRRKRFMPKV---VSEGKPKRPRKPAELPKVVVEGKPKRPR 80

Db 117 nliecnldqelklfvshrsharfspvgqklhrserfxetvvlqpfllmegssgtevr 176

Qy 81 KAATOEKVKSK-----ETGSAKKKLKE-----SATKKPANVG 113

Db 177 lmit-daarkhllvtgqcfentgelilqsgsfqnfieftdqeigellstthpanka 235

Qy 114 DMSNKSPVTLKSCRKALNFDLENPGD-ARQOSESEIVQNSGANSFSEIRD----- 165

Db 236 sltlfcpe---egdwksnldrhnlqdfiniklnsasilpemelglsefteylsesveps 292

Qy 166 -----AIGGTN-----GS-----FL 175

Db 293 pfdileptsggflklskpcyifpggrgdsalfavngfnmlnggserksctfwklrlh 352

Qy 176 DSVSQIDKT---NGLGAMNQPLEVSMGNQPDKLTGAKLARQQDPDLLTTRNOCCQPPVA 231

Db 353 drvdslilthlgddnlpnsmldqrkaeleeedsgstnsdwmknlsdpdglgvflnv 412

Qy 232 TQNTQFPMENQQAWLQMKNLIGFPFGNQQRPMTRNQQPCL-AMGN----- 277

Db 413 penlknpepn---ikmrslieacftlqylnklsmppeplfrsvngtldpvlilqkmvg 468

Qy 278 -QOPWYLIGTPRPA--LVSGNQQLGG-----POGNRPPIFLNHQITCL-----P 317

Db 469 gklemyvlnpvksskemqymqwtgtnkkaefilpngqevdipilayltsvssliwhp 528

Qy 318 AGNQLYGSPTDMHOLVMSTGGQ-QHGLL-----IKN-----OQPGSL---IRGOQPCVPLI 364

Db 529 a-----npaekiirvifpgnstqynileglekikhldfkqlatqkdlgtgvp-tpvv 581

Qy 365 DQQPATPKGFTHLNQM-----VATSMSSPGLRPHSQSQVP--TTYLHVESVSRILN 413

Db 582 kq-----tklqgradsreslkpaakplpsksvrkeskeetpevtkvnhve----- 626

Qy 414 GTTGTCQRSRAPAYDSLOQDTHQGNKYILSHEISNGNGCKKALPQNSLSLPTIMAKLEE 473

Db 627 -----kppkveskekvmvkkdkpvkte--tkpsvtekevpskee-pspvkaevae 674

Qy 474 RGSROYHRANGOTEKHDLNLAQQAQSO-----DVERHNSSTCVEYLDAAKTKIQKV 527

Db 675 qatdvkpaakektvkketvkpedkkeekpkveakkedktpikkeekpkkeevkke 734

Qy 528 VQENLHGMPPEVIEDDPTDCARKGNKTASISKASGKNSSPVKTKAEKCIIVPTTPA 587

Db 735 vkkeikk-----eekpekkettpkpekketpkvekkkevkk-eeke-----pkkel 783

Qy 588 KKGRAGRKKSVPFAHASEIOLMPTPP-----KTPLSRSKPKGKR-KSIQDSGKAR 639

Db 784 kklpkdaksstplseakkpaalkpvkpkkesvkkdsaaagkpkkgkikvikkekaa 843

Qy 640 GPSGELLQDSIAEIIYRMQNLVLDKEREQFONAMWLYKGDGALVPVESKRRPRPRVD 699

Db 844 eaavaavtqattaavaamaaaglaaipakelaerslmspedltckdfeelkae----evd 900

Qy 700 IDDETTRIWNLMKGDEKDEKDKKKEK-----WVEEERRVFERGRADSFIARMHLVQG 755

Db 901 vtddikpqlleli-----edeeklketpevaeaylqkerevtkgpaesdegittteg 952

Qy	756	D-----RRFSPWKGSVVDSVIGVFLPTQNVDHLLSSGSAFMSLAARFPKPLUSSREDERN	808
Db	953	egeceqtpeelepvekqvddi-----ekfedegagf-eesestgdyeeek	996
Qy	809	VRSVVEDPES-----CILNLNEIPSWOEK-----VQHPSDMEVSQVDSGSKEQL	853
Db	997	aeteeaepeedgehvcvsaaskhptedeesaakaedayirekresvasgdrraedem-	1055
Qy	854	RDCSNSGIERNFLEKSIGNLEEVLSS---QDSFDP-----AIFOSCGRVSGCS	900
Db	1056	---deaiek-geaqseeeeadeekaareeeyepekeameadyvmavvdkaaaagae	1110
Qy	901	-----CSKSDAEFTTRCETKVTSVGTSQSVOVOTSPNLSDCEICLOGERPHLY	947
Db	1111	eqvgflttptkqlgadsprepasishdtlpggseseatasdeenre-----dqpeef	1164
Qy	948	EGSG-----DYOKOETTNAQAKKPDLEKTWNWKSVCFCGOPR	984
Db	1165	tatsgytqstielisseptmdenstprdvmsdetunteespsqfv-----	1211
Qy	985	NDTNWQTTPSSSYEOCATROPHYLDIEDF-----GMQGEGLGYSNMWSISPRYDRVKNK	1037
Db	1212	nltkye---sslysqesykpadvtpingfsegsktdatdkgdynasaastisppssmeedk	1268
Qy	1038	NVPRPRFGQSVPRFTGGIITSPHPHELPGMLGSSSVAQEHODDTQHNOQDEMKNAS	1097
Db	1269	----fsr--saldaycsevkastt-----ldikdisasvssekkspskspspspps	1316
Qy	1098	HLOKTEFL-----DLLNSSECUTROSS--TKONTIDGCL-PRDTAEDVVDVP-	1141
Db	1317	plektplgersvnfstlptneikvsaaevapvspetqevveehcaspedtkle-vvsps	1375
Qy	1142	----JSNNSSLONILVESNNSKKEOTAVE-----YKETNATILREMKGTL	1182
Db	1376	qsvtgsaghtpyyqgstdekshnlpteieviekkppvypvsfetsdakdenerasvpmdepv	1435
Qy	1183	ADGKKP-----TSQWDSL-----RKDVNEGREGROERKNKMWDSIDYEAI	1221
Db	1436	pdaespiekvlslrppligesayesfiladdkasrgaespfEEKsgkqgsdp----	1491
Qy	1222	RRASISEIAI-----KERGMNMALAVRIKDFLERVHDHGIGDLEWLRESPPDKAK	1274
Db	1492	--qvpvsemstslsyqdkqegkstdfapikedfgkektd----dreamssqpa----	1540
Qy	1275	DYLLSIRGLKLSVECVRLTLTHNL-----APFVDNTV-----G	1308
Db	1541	-lalderklgdvsptqdsvsqfsgfkcdkmsisegtvsdksaatpvdgvaedytsymeg	1599
Qy	1309	RIAVRMCWPLOLPPESLQLHLELVPVLESQKFLW-----PRCLKLDQRTLVELHYQL	1363
Db	1600	vavstasvatssfpe-----pttdcvpslhaevgshstveddsilsvvvqtp	1649
Qy	1364	ITFKVFCTKSRNCNACPMRGECRFASFAYASARLALPAPEERS-LTSATIPVPPSEFP	1422
Db	1650	ttfqetempskeecpr-pmnsipdpfspktaksr-ltpvqdhreqsgsmisiefqges-p	1705
Qy	1423	PVAIPM-IELPPLERKSLAGSAPSRENCEPIIEBPSPGOECTEITESDIEDAYYNEDP	1481
Db	1706	eqslamdfrsqsdhptvgagvlhten-----gptevdyspdmqdsash--	1752
Qy	1482	DEIPTTKLNTEOQGMTLREHMERNMELOEGDMSKAL----VALHPTTTSTIPTPKLNISR	1537
Db	1753	-kipp-----meepsytqndilselivsqveaspstsahtp-----sq	1791
Qy	1538	LRYEHQVVELPDSHRLLDGMDXREPDPSPYLALTWTGTETANSQPPEQKCGGKASKMK	1597
Db	1792	iasplqedtl-----sdvapprdmslja-----	1814
Qy	1598	CFDETSECNCREANSQTVRGTLTLP---CRTAMRGSFPLNGTYFO-----VNELFAD	1648
Db	1815	-----slctekvqilegeklspskdsisltpressplyptsfdstsavkektat	1864
Qy	1649	-HESSLKPIDVPRDWIDLPRRTVIVFGTSV--TSITFRGLSTEIO-----FCF	1693

Db	1865	chessppidaaasepygf-rasvldftmqhhlainrdistpglekdgsgktpgdfsyay	1923
Qy	1694	WKGFCVCRGFEQKTRAP	1710
Db	1924	qkp-----eetrsp	1933
RESULT 10			
ABG17148			
ID	ABG17148	standard; Protein; 2759	AA.
XX	AC	ABG17148;	
XX	DT	18-FEB-2002	(first entry)
XX	DE	Novel human diagnostic protein #17139.	
XX	XX	Human; chromosome mapping; gene mapping; gene therapy; forensic;	
KW	XX	food supplement; medical imaging; diagnostic; genetic disorder.	
KW	XX	Homo sapiens.	
OS	XX	W0200175067-A2.	
XX	PN	11-OCT-2001.	
XX	PD	30-MAR-2001; 2001WO-US08631.	
XX	PF	31-MAR-2000; 2000US-0540217.	
XX	PR	23-AUG-2000; 2000US-0649167.	
XX	PR	(HYSE-) HYSEQ INC.	
XX	PA	Drmanac RT, Liu C, Tang YT;	
XX	PI	WPI; 2001-639362/73.	
XX	DR	N-PSDB; AAS81335.	
XX	DR	New isolated polynucleotide and encoded polypeptides, useful in	
PT	PT	diagnostics, forensics, gene mapping, identification of mutations	
PT	PT	responsible for genetic disorders or other traits and to assess	
PT	PT	biodiversity	
XX	XX	Claim 20; SEQ ID No 47507; 103pp; English.	
XX	PS	The invention relates to isolated polynucleotide (I) and	
XX	CC	polypeptide (II) sequences. (I) is useful as hybridisation probes,	
XX	CC	polymerase chain reaction (PCR) primers, oligomers, and for chromosomal	
XX	CC	and gene mapping, and in recombinant production of (II). The	
XX	CC	polynucleotides are also used in diagnostics as expressed sequence	
XX	CC	for identifying expressed genes. (I) is useful in gene therapy techniq	
XX	CC	to restore normal activity of (II) or to treat disease states involv	
XX	CC	(II). (II) is useful for generating antibodies against it, detecting o	
XX	CC	quantitating a polypeptide in tissue, as molecular weight markers and	
XX	CC	a food supplement. (II) and its binding partners are useful in medical	
XX	CC	imaging of sites expressing (II). (I) and (II) are useful for treating	
XX	CC	disorders involving aberrant protein expression or biological activi	
XX	CC	The polypeptide and polynucleotide sequences have applications in	
XX	CC	diagnostics, forensics, gene mapping, identification of mutations	
XX	CC	responsible for genetic disorders or other traits to assess biodiversi	
XX	CC	and to produce other types of data and products dependent on DNA and	
XX	CC	amino acid sequences. ABG00010-ABG30377 represent novel human	
XX	CC	diagnostic amino acid sequences of the invention.	
XX	CC	Note: The sequence data for this patent did not appear in the printed	
XX	CC	specification, but was obtained in electronic format directly from WIP	
XX	CC	at ftp.wipo.int/pub/published_pct_sequences.	
XX	CC	Sequence	2759
XX	CC	Sequence	2759

Matches 331; Conservative 221; Mismatches 625; Indels 649; Gaps 82;		
Qy	52	VEGKPKRP-----BKPAELPKVVEGKPKRKAATQEKV-----88
Db	646	lqggsrvptlqgphltnkspasssfqggspsasptvntqgqmgprppqnnlppq 705
Qy	89	-----KSKETG-SAKKKNLKESATKKPANV 112
Db	706	fqqpvsppgnpmvqgnvppnfmvmqagqppnqgqslhpglgmpkrlpplgfsagqanp 765
Qy	113	GDMNKGPEVTLKSCRA-----LNFLENPGDARGDSESEIVQNSGANFSE-----162
Db	766	nfmqgvpsttattcpngsagqlqanqnvhagggagqppqng-mqvsghppnmmqslm 824
Qy	163	IRDAIGTGSGFLDSVQIDKTNGLGAMNQ-PLEVSMGNQPKLSTGAKLARDQ 216
Db	825	gihnmnnqagtsq-----vpqvnlnnmqgqgqgppslqmgmhqqlvpsqgqmvqgq- 878
Qy	217	PDLLTRNQOQCFPVATQNTOPFMENOQAWLQMKNLIGFP-----FGNQQPRMTIRNOQPC 272
Db	879	-----gtlnpqnmilstraqlmpqgqmmvnpqslgspqgmt-----917
Qy	273	LAMGNQPMYLIQTPRALVSGNQQLGCP-----QGNKRPIELNH 312
Db	918	-----ppkqmlsqgqgmaphnqmgppqgqvlqgnpmieqimtnmqmgnkqfntqn 971
Qy	313	QT-CLPAGNQLYGSPT-DMHOLVMSTGQOQHGLLKKQOP-----GSLIRGQ 358
Db	972	qsnvmpgaqlmrtpbmgmnmvqfvgmsqgmipqgqgvnnpspsqvmglqgvlrppg 1031
Qy	359	PCVPLIDQQATPKGFTHLNQWATSMSSGFLRPHSOSQVPTTYLHVES-----408
Db	1032	ps-phmaqqhqdpa--ttannndvslsqmmpdvsigqtnmvpv---hvqamqnsasguh 1085
Qy	409	-----SRLTNGTGTQCSRAPAY-----DSLQDDHQ-----N 438
Db	1086	sghmsfnafsgapnqmscgqn--pgfpvkvdxltsppllvnlqdsiasghfgvnn 1143
Qy	439	KYILSHETISNGCKKALP-----QNSLPTPIMAKLEEARSGSKROYHRAMQOTEKH 490
Db	1144	k-----qmntnankpkkkpprkkxnsqqlntptripagleead-----1183
Qy	491	DLNLAQIAQSDVERHNSCTCV-EYLDAAKTKIQVQBNLHGMPEVIEEDPTDG 549
Db	1184	-----qpplpgeqnlndnsgkplbefanrppgysqpvqdrplqgmppqlmqhvapp- 1236
Qy	550	ARKGKNTASISKGASKGNSPVKTAKEKCIVPKTAKGGRAGKSKSVPPAHASEIQL 609
Db	1237	-----ppqpqq-----pqqlp---qq 1251
Qy	610	WQPTPKTPLSRKPKGKRKSI-----QDSGKARGPSGELLQDQSI AEIIRMQNLYLG 664
Db	1252	qppppsqpsqgqqqqqqqqmmmmmmmqgqpksvrlp-----vsqnvhpgrplnp 1302
Qy	665	DKEREQONA-----MWLYKGDGALVPYESKRRKPRKVDIDDTTRINLLMCKGDEKE 719
Db	1303	dsqtrmpmqdgsqsvpmvmslqgpasvpbspdkgrmpmv-----ntplgsnsrkm 1351
Qy	720	GDEKDKKKEWEEERRVFRGRADSFIRM-----HLVQGD 756
Db	1352	vygespqp-----ssslaeamaslpeasgsaepsvpggpnmpshvlpq 1397
Qy	757	RRF---SPWKG-SVVDYSVIGVFLTQNSDHLSSS---AFMSLAARFPFKLSSSREDER-N 808
Db	1398	nqlumtgpkpqpslataqatqgppvnslpssghhfpnvaa-ptqtsrptknras 1455
Qy	809	VRSVVVEPPECILNLNLEIPSWQEKVQHPSDMEVSGVDSGKEQLRDCNSGI---ERENF 866
Db	1456	prpyyptpn-----nrppste-----pseisl-----sperl-nastaglfppqini 1497
Qy	867	LEKSIONLEEVLSQDSFDPALFOSGCRVGS-CSCSKSDAEFPTT-RCETKTVSGTSQS 924
Db	1498	plpprpnlnrgf--dqglnpnttklqgqapsnlmtmopsnfatpqthkldsvvvnsgkqs 1555

Qy	925	VOTGSPNLSDICLOGNERPHLYEGSDVQOKETTINVAOKKPDLEKTMNMKDSVCFQOPR 984
Db	1556	-----nsgatkraspsnrrssp-----gssr 1577
Qy	985	NDTNWOTTPSSSYEQCATROPVHVDIEDFCMQGEGGLGYSMWSISPRVDRVKNKNVPRPF 1044
Db	1578	-----ktlpsp-----grqn-----skapkltlasqnaa---1 1603
Qy	1045	ROGGSVPREFTGIIIPSTPHELPGMLSGSSSAVQEHQDDTQHNOQDEMKNASHLQKTF 1104
Db	1604	lqnvlpnrvlvsptplanppvpsfpnnsg-----1634
Qy	1105	DLNLSSECIUTROSSYKQNTIDGCLPRDRTAEDVVDPLSNSSL---QNTLVESNSNKE 1161
Db	1635	-----lnpqnstvsvaavgvved-----nkeslnvpqdsdcqsqrke 1674
Qy	1162	QTAVEYKETNATILREMGTLADGKPTQOWDSLKRDVEGNEGRQ-----ERNKNMDSID 1217
Db	1675	qvnlelkavpa---qevkmvpe-----dqskdgqpsdpnkipsveenklvs---1720
Qy	1218	YEAIRRASISEIAIKERGMNMMLAVRIKDFLERIVKDHGGIDLEWLRSPDPKAKDYL 1277
Db	1721	-pamreapt-sqlldnsgapn-----tikppgltdle---vtp-1757
Qy	1278	LSIRGLGLKSVCEVRLTLHLNLAFFVDNTNYGRIAVRMGWVLPQLPESLQLHLELYPVL 1337
Db	1758	-vvsgeedkkaasvip--tlqdlssskpsns-----lnlphsnelcsslvhpel 1803
Qy	1338	ESIQKFLWPLRLCKLDQRTLYE-----LHYQLITFGVCTKSRPNCNACPMRGECRHA 1391
Db	1804	sevsnnvapslppvmrpsvsssisltplppnqit---vftv-snp-----1845
Qy	1392	SAYASARLALPAPEERSLTSATIPVPPESPVAIPMIELPLPLEKSLASGAPSNRENCE 1451
Db	1846	tsaantaaalptlqsalmtvtvmpnag-----skmvsqggaagsnar 1891
Qy	1452	PIIEEPASPOQECTETESDIEDAYNEDDEIPTIKLNIE-----QFGMTUREHMERH 1505
Db	1892	pqfitp------insssliqvmkgspstipaapltnsglmpsvavvgplhipqn 1944
Qy	1506	MELQEGDMSKALVALHPTTTSITPKLKN-----ISRLRTEHOVYELP-----DSHRL 1553
Db	1945	ik-----fssapvpnnalsss-papnliqtgrlvissratcpvlpspctssvvpvshpp 1998
Qy	1554	LDGMDKREPDPSPYL-----LAIWTPGETANSQAQPPQKCG-GK 1592
Db	1999	vqvkelnpdeaspqvntsadqntlpsqsttmvpsplltnspsgsgnrrrpsvsskkgk 2058
Qy	1593	AS--GKMCFDCTCE-CNSLREANSQ 1615
Db	2059	vdkgilltkackkvtslekgeeq 2084
RESULT 11		
AAU14697		
ID	AAU14697	standard; Protein; 5447 AA.
XX	AAU14697;	
XX	AAU14697;	
DT	24-OCT-2001	(first entry)
XX	Novel bone marrow polypeptide #96.	
DE	Bone marrow; diagnostic; therapeutic; gene therapy; antigenic;	
XX	haematopoiesis; myeloid; lymph cell disorder; tissue regeneration;	
KW	wound healing; nutritional supplement; immune disorder;	
KW	severe combined immunodeficiency; SCID.	
OS	Homo sapiens.	
XX	WO200157187-A2.	
PN		
XX		

PD 09-AUG-2001.
XX 05-FEB-2001; 2001WO-US03782.
XX 03-FEB-2000; 2000US-0496914.
PR 20-JUN-2000; 2000US-0598075.
PR 19-JUL-2000; 2000US-0820325.
PR 30-NOV-2000; 2000US-0250683.
XX (HYSE-) HYSEQ INC.
XX Ford JE, Boyle BJ, Tang YT, Liu C, Asundi V, Zhou P, Xue AJ;
PI Ren F, Drmanac RT;
XX WPI; 2001-488875/53.
DR N-PSDB; AAS23002.
XX Nucleic acids encoding bone marrow polypeptides, useful in diagnostic
PT and gene therapy -
XX Claim 10; Page 124-127; 392pp; English.
XX AAU14602-AAU14794 represent novel bone marrow polypeptides of the
CC invention. The proteins and corresponding coding sequences may be used
CC in the prevention, diagnosis and treatment of diseases associated with
CC inappropriate bone marrow polypeptide expression. For example, to treat
CC disorders associated with decreased expression by rectifying mutations
CC or deletions in a patient's genome that affect the activity of the
CC polypeptides by expressing inactive proteins or to supplement the
CC patient's own production of the polypeptide. Additionally, the nucleic
CC acids may be used to produce the polypeptides, by inserting the nucleic
CC acids into a host cell and culturing the cell to express the protein.
CC The nucleic acid and its complementary sequences may also be used as DNA
CC probes in diagnostic assays to detect and quantitate the presence of
CC similar nucleic acid sequences in samples, and therefore which patients
CC may be in need of restorative therapy. The proteins may also be used as
CC antigens in the production of antibodies against bone marrow proteins
CC and in assays to identify modulators of their expression and activity.
CC The anti-bone marrow protein antibodies and antagonists may also be used
CC to down regulate expression and activity. The antibodies may also be used
CC as diagnostic agents for detecting the presence of the protein in samples
CC (e.g. by enzyme linked immunosorbant assay (ELISA)). The proteins
CC may be used to regulate haematopoiesis activity, and consequently in the
CC treatment of myeloid or lymph cell disorders; in tissue regeneration,
CC such as wound healing; as a nutritional supplement; and in treatment of
CC immune disorders such as severe combined immunodeficiency (SCID).
XX Sequence 5447 AA;
SQ
Query Match 2.4%; Score 222.5; DB 22; Length 5447;
Best Local Similarity 17.4%; Pred. No. 1.1e-06;
Matches 322; Conservative 298; Mismatches 701; Indels 527; Gaps 78;
QY 16 TEQNDGSRQDVFLEDLNKTPOQKSKRRKPKMPKV-----VVEGKPKRPRKPAELPKV 69
DB 1594 ttqgdlsalqknqsdI-kdiqddiqnratfsatvkvkdegfmeenqtklspreital--- 1649
QY 70 VVEGKPKRPRKPAELPKVSKETGSAKKNLKESATKKPANVGDMNSKSEVILKSCRK 129
DB 1650 -----reklhqakeqyealqeetrvaqke-leeavt-----s 1680
QY 130 ALNFDLENPGDARGDSESEIVQNSGANSFSEIRDAIGTNGSFLDSVQIDKTNLIGA 189
DB 1681 alqetekskaak-----elaenkkkidaIldwvtvsgsggqllnlpgmeqlsgasl 1734
QY 190 MNQPLEVSMG-----NQ-PDKLTGAKIARQOQPDLLTRNQOCQPPVATQNTQFFPENOQA 244
DB 1735 ekgaIdtdgymgvnqapeklkdcemmkarhqell--sqgnqfilatqsaqlfdqgh 1792
QY 245 WL-----OMKNQLIGFFPGNQOPRWTIRNQOPLAMGNQOPWYLIGTPRALVS----- 293
DB 1793 nltpeeqmIqqkIq-----elkeqysylaqseaelkqvltlqdelqkflqdhk 1842

QY 294 -----GNQOLGGPQGNKRPIFLNHQTCPLPAGNQLYGSPTDMHQLVMSTGQO 340
DB 1843 efeswlersekelenmhkggsspetlpsllkrq-----gs-----fsedvishkgdl 1889
QY 341 HGLLIKNQO-----PGSLIRGOQPCV---PLIDQOPATPKGFT-----HLNOMVA 382
DB 1890 rfvtisgkvldmensfkegkepeignlvkdklkdaterytalhskctrlgshlnmlilg 1949
QY 383 TSMSSPGLRPHSQSQVPTTYLHVESVSRIILGTGTQCRAPAYDSLOQDIHQNKYIL 442
DB 1950 -----qyhqfqsadsIqamwqacean-----veklldstaasdpvvl 1987
QY 443 SHEISNGCGCKALPONSSLPPTIMAK-----LEEARSGKROYHRAMGQTEKHDLNLAAQOI 498
DB 1988 qegIattkqlqeelaeh-qypveklkvqardimeiegepapdhrhvqettdsIlshfqsI 2046
QY 499 AQSDQVERHNSCTVEYLDAAKTKIOKV-----VQENLHGMPPEVIEIEDPTDQARK 552
DB 2047 s-----yslaerssllqkaIaqsqsvqdslesllqslsigeveqn-----le 2086
QY 553 GKNTASISKASKGNSSPVKKTAEKKECIVPKTPAKKGRAGRKKSVPVPAHASEIQLWQP 612
DB 2087 gkvssIsgvi-----qeatnmklkqdiarqkssI--eatremvtrfme 2131
QY 613 TPKTPTLSRSKPKGKRKSIODSKARGPSGELLQCDQSIAEIYRMQNLVGLDKEREQEQ 672
DB 2132 tadsttaa-----vlgqlaevsqrfegIclqgqek----- 2162
QY 673 NAMVLYKGDGALVPYESKKRPPKPRKPIDDETT-RINWLLMGKDEKEGDEEKDKKKEK 731
DB 2163 -----essllklIpqaemfehlsgklqgfmenksrmlasngnqpdqdlthf 2207
QY 732 WEERRRVRGRADSFIARMHLVQGRDRFPWKGVSDVSVIGVFLTONVSDHLSSAFMSL 791
DB 2208 fqq-----iqelnlemedqgen-----ldtl-----ehlvtelsscglfald 2243
QY 792 AARPPPKLSSSRREDNRVSVVVE-----DPGECILNLNE-----IPSWOEKVO---HPS 838
DB 2244 lcqhqrdrvqnIrkdfteIqktvkeredasscqqldefrklvrtfkIkwIketegsIpt 2303
QY 839 DMEVSGVD-----SGSKQELRDCNSG--IERFNFLKSKTONLEEEVLSQD-----SFD 886
DB 2304 etmsakeIekIqtehlkllldwaskgtlveelnlykgtslenIteIapdsqgktgsil 2363
QY 887 PAIFQSCGRVG---SC-----SCSKSDAEFP-----TTRCET---KT 917
DB 2364 psvgssvgsvngvyhtckdlIteIqcdmsdvnlkyeklgglvherqesIqalInrmeevhke 2423
QY 918 VSGTSQSVQT-----GSPNLSDICLOGNER-----PHLYEGSGDVQ---KQETT 960
DB 2424 ansvlqlweskeevlksmdamssptkktetvkaqesnkafIaeIeqnspkIqkvkealag 2483
QY 961 VAQKKPDLKTMNWKSDVCFQPRNDTNW-----QTFPSSSYEQC-----ATR 1003
DB 2484 Ilvtypnsqaeenwkki-----qeelnrweratevtvarqrleesashIacfqaaesql 2539
QY 1004 QPHVLDTE-DFGMOGEGILGYSWMSISPRVDRVKNKNVPRFRFGQSGVPREFTOGIIPST 1062
DB 2540 qpwlmeIkelmmgvIgp-----IsidpnmIoaqkqv--qfmIkefearrqheqlneaa 2591
QY 1063 PHELPGMG-LSGSSSAVOEHQDDTQHNOQDMMKNKASHLQKTFDLNLSSECIITRQSTTK 1121
DB 2592 qgIltpgvdvsIstsqvqke-----lqslnqkwvltcdkInsr 2629
QY 1122 QNITDGLCPDRTRAEVDVPLSNSSL--QNILVESNSSKEQTAVYKTNATILREMK 1179
DB 2630 ssqIdqalvktstgqellqldsekvragrIvqsaIstqpeavkqqlseIseIradle 2689
QY 1180 GTLADGKKPTSQWDSLRKDYEGNEGROERNKNMDSIDYEAIRRASISEISEAIKESGMN 1239
DB 2690 qldhevkeaqIcdelsvll-geqylkdelkrlIetv-----alplggle 2733

QY 1240 NMLAVRIKDFLERIVKHGIDGL-----EWLRESPPDKAKDYLLSIRGLKLS-----VE 1289
Db 2734 dilaadrinriqaalastqfqcmfdeirtrwddkqsqakncpisaklerlqsglqence 2793
QY 1290 CVRLTLHNLAFVDTNWGRIAVMGWVLPQIPESIQHLLLELPVLESIQHLPRLC 1349
Db 2794 fqklnhsgsyevivaegeslllls---vppgeekrtlqnlvelknhweelskktadrqs 2851
QY 1350 KL----DQRTLYELHYOLIFGKVFCTKSRPNCNACPMRGECRHFASAYASARLALPAPE 1405
Db 2852 rlkcdmqakqyqhve-----dlvpwiedck---akmselrvtlid--- 2889
QY 1406 ERSLSATIPVPPSEFPFPAIMIELPLEKLS-LASGAPSN-----RENCEPIIEEPASP 1460
Db 2890 -----pvqllessllrskamlinevekrslleilnsaadi 2923
QY 1461 QOCTETESDIED--AYYNEDDP-----EPTTKLNIEQFGMTLREHME--RNWELQ-EG 1511
Db 2924 linsseadedgirdkaginqmmdavteelqaktgsleemtqrlrefqesfkniekveg 2983
QY 1512 DMSKALV--ALHPTTTSIPTPKLNISRLRTEHOVYELPDS-----HRLLDGMDKREPD- 1563
Db 2984 akhgleifdalgsqacs-----knleklraqgevlqalepqvdylnrfcgivedapdg 3038
QY 1564 -DPSFYLLAIWTPGETANSAPPEQKCGKASGKCMCFDETCSECSLREANSOTVRGTL 1622
Db 3039 sdasqll-----hqaevaagqe-----flevkqrwns-----gcvm 3068
QY 1623 IPCRTMRGSPFLNGTFVYNELFADHESLSKPIDVPRDWLWDLPR 1670
Db 3069 menkleglqgf-----hcrvremf-----sqliadldelgmgairdt 3107

RESULT 12
ABB67502
ID ABB67502 standard; Protein; 3257 AA.
XX
AC ABB67502;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 29298.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
XX
KW pharmaceutical.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
XX
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
XX
DR N-PSDB; ABL11605.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
XX
PT genes from Drosophila and for elucidating cell signalling and cell-cell
XX
PT interactions -
XX
PS Disclosure; SEQ ID NO 29298; 2lpp + Sequence Listing; English.
XX
XX
XX The invention relates to an isolated nucleic acid detection reagent
XX
CC capable of detecting 1000 or more genes from Drosophila. The invention is
XX
CC useful in developmental biology and in elucidating cell signalling and

cell-cell interactions in higher eukaryotes for the development of
insecticides, therapeutics and pharmaceutical drugs. The invention
discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
sequences (ABL01840-ABL16175) and the encoded proteins
(ABB57737-ABB72072).
The sequence data for this patent did not form part of the printed
specification, but was obtained in electronic format directly from WIPO
at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 3257 AA;

Query Match 2.4%; Score 222; DB 22; Length 3257;
Best Local Similarity 17.6%; Pred. No. 5.2e-07;
Matches 349; Conservative 273; Mismatches 775; Indels 586; Gaps 77;

QY 7 SSAVNATEATEQNDGSRQDVLFEFLNKTPOOKPSKRKRKF-----MPKVVE 53
Db 798 sskdeppaeenlpapdqdpie--qgktpvakngqhdkehneapkaeslsvsdipssvt 855
QY 54 GKPKRPRKPAELPK--VVVEGPKRPRKAATQEVKSKETSGAKKKNLKESATKK-- 108
Db 856 pskrnhsspantpkkskealealqssvrralrsdk-----atpqniresrskrtl 907
QY 109 -----PANVGDMSKNKSP---EVTLSKCRKAL-----NFDLEN 137
Db 908 teitllmddtmrrssprlgrspaeshsherspmekktvtvsklakdlitidkekeitel 967
QY 138 PGARQGD-----SSEIVQNSSGANSFSEIRDAIG-GTNGSFSDVSQIDKTNG 186
Db 968 lpadsetkdvkikttaasdtltdenpssttemkklgkplkakkmartetevkka 1027
QY 187 LGAMNQ--PLEVSMGNQPKLSTGAKLARDQOQDLTLRNQO--COFPVATQTFPMENQ 242
Db 1028 iadsnedipsifskveehltseseqdekeellcpkqidctndlqgstaietde 1087
QY 243 QAWLQMKNLIGFPFGNQ-----OPRMTIRNQO--PCL 273
Db 1088 qveekrsnrirkrirnekkftetdtlsdhldakkaenasleismrpkctietqgsdpt 1147
QY 274 AMG-----NQOPMYLIGTPR-----PALVSNQOQLGGPQGNKRPILFNH-----QT 314
Db 1148 akknrsgrlirkeksvinaaksekdkpsais-----qsterqllnepsdkdkk 1199
QY 315 CLPAGN--QLYGSPTDMHQLVMSTGGQOHLTKNQPGSLIRGQQPCVPLIDQOPATPK 372
Db 1200 teqsgnkkavvgpldkttetsstn-----iidkkneshfamsqpsdrlnqkesa-- 1250
QY 373 GFTHLNQWVATSM-----SSPGLRPHSQSQVPTTYLVHVES----- 407
Db 1251 -ftklssissppkkimkdqdkldalskgsdnptirdtgedsrqtdkhhgendckheed 1309
QY 408 ---VSRILNGTGTGCRSRAPAYVDSLOQDI-----HQCNGKIVLSHEI 446
Db 1310 ssklikanidetkssekdaepiskdssqdsakprlskpkprnkrkknekpkndsiaesdi 1369
QY 447 SNG-----NGCKKALPQNSSLPTPIAKLEEARSGKR--QYHRANGQT 487
Db 1370 eggfqvtetvqatcstpsesnkkdmvksdetneepnlseteigrirkrqgafhiempkd 1429
QY 488 EKH-----DLNLAQOIAQSQDVERHNSSTCV-----EYLDAAK-KTKI----- 524
Db 1430 dlhitpqnqnsiagvnfekqvpipeve---sdtplmkptktylmctknktsllsase 1486
QY 525 -----QKVQVQNLHGMPEVIEIDDDPTDGARKGNKTASISKGASKGNSSPVKKTAE 576
Db 1487 dpdivleqqlittsksgdnpdlannletstqdpke-hefsdqftdndsiipscbk 1545
QY 577 KEKCIIVPKTPAKKGRAGRKKSVPV-----PAHASEI 607
Db 1546 ksqivftptktssdqtknsfitpnrspskksrwnvskeakrlndnsfeesnaasessaskv 1605
QY 608 QLWQFTP-----PKTPLSRSPKKGKRSIQDSGKARGPSGELLQDSIAEIIY 656

```
Db 1506 qkelrtpascrkrlrvlkrtrpts-slpt-nsrksifktpaksrkrltkil-esmektps 1662
Qy 657 RMQNLVYG----DKREQEQNAMVLYRGDQALVPYESKRPKRPKVIDDIDDETRINLLM 712
Db 1663 repvs1gevnpsdpvaesav1hesdrlesei-----pneevfedt----- 1708
Qy 713 KGDEKGEDEKDKKKKEWEEERVRFRGRADSF1ARMHLVQGDRRFPSPWKSGVSDSVIG 772
Db 1709 ---eeaaedtdnk1kkkeddhelev-----ndicaasknpitddstkdassnkstdsdv- 1760
Qy 773 VFLQNVSDHLSSAFMSLAARFPKPLSSREDERNVRSVVVEDPEGCILNLEIPSWQE 832
Db 1761 ---lqetdelslninatqgdtpikelteeevpnukt-ved----- 1799
Qy 833 KVQHPDMEVSGVDSGSGKEQIQCNSNGIERFNFLEKSIQNLDEEVLs---SODSFDPAI 889
Db 1800 -----eskkqelkd-----lepdnaaleedastakaaeamdyi 1835
Qy 890 FQSGRVGSCSCSKSDAEFTTTRCETKTVSGTSQSVOTGSPNLSDEICLQGNERP----- 944
Db 1836 ke-----ksnvksvlaepet-dvtddeelaqspipnssettsvtdppestsv 1883
Qy 945 ---HLYEGSDVQKQETTNVAOKKPDLEKTMNWKDSVCFQGNRDTNWQTPSSSYEQCA 1001
Db 1884 vkslrkreadssqpdeaaakrkqrdveksltgk-----qvkpa----- 1924
Qy 1002 TROPHVLIDIEFGMQGEGGLGYSMWSISPRVDRVKNKNVPRFRPQ-----GGSVPREFT 1055
Db 1925 -rrqlaeeve-----rpslrksteseakstvggkyisignetsms 1968
Qy 1056 GOIIPSTPHELGMLSGSSSAVQE--HODDTHNOQDEMKN-----ASHLQKTFLD 1105
Db 1969 tapiretnreaastpsarksavqeahvettkhilpgpgkllhdspeaevkqpmvq 2038
Qy 1106 LLNSSECLTRQS-----STKQNTDCCLPRDRTAEDVVDPLSNSSL----- 1148
Db 2029 tllsstls1qkpsltdgsp1k1rks1kks1adengdgs1fsssvlnknts1vapr 2088
Qy 1149 QN1LVE--SNSNKEQTAVEYKETNAT1REMKGTLADGKKPTSQWDSLRKRDVEGNGRQ 1206
Db 2089 vn1svllqsktdqvetaasasetp1ltkktkstktp-----egnkkte 2137
Qy 1207 ERNKNMDSIDY-----EATRASI3EISEAIKERGMNNMLAVRIKDFLERIVKDHGGI 1260
Db 2138 skkkslvqgpmktqkseeavsgpki--lnkylksetessrktvtstgrkqig----- 2190
Qy 1261 DLEWLRESPPDKADYLLSIRGLKSVCEVRLTLNHLAFPPVDTNVRTAVRMGWVPLQ 1320
Db 2191 -lev1kpesrkseesi--veaisrkkqsgvqriskidgrkseqtslpqpdvkssetalk 2247
Qy 1321 -PLPESLQLHLELYPPVLESTQKFLWPLRCKLDQRTLYELHYQLITTFKVFCTKSREN 1379
Db 2248 aalpkete-----fpvqda-----eiekmskgrghqnavkntkteqpskpkte 2291
Qy 1380 ACPMRGECR-----HFASAYASARIALPAPERS 1408
Db 2292 vrs1qaeaatelmdsmdsqsdvsdiratfpesqg1fnvpghmtraissnrs1aptptms 2351
Qy 1409 LTSATIPVPPSPFPVAIPMTLEPLPLEKSLASGAPNRENCEPIIEEPASPGQECTEI- 1467
Db 2352 dsqrn--askerftcpvsdq--kkp1resqtlskrarrgrnqnp1vskrkageadgtavi 2407
Qy 1468 -----TESDIEDAYNED-----PDEIP-----T 1486
Db 2408 npkprpremedehpqqdnhvqesaafa1p1v1kitaassv1pqvvrstgntvpqns1prk1c 2467
Qy 1487 IKLNIEQFGMTLRHEMRNMELQEGD---MSKALVALHPTTTSIPTPLKNI---SRLRTE 1541
Db 2468 vk1nrrpynkwlrstqernee-qegsnv1sl1pl1getsetdaaesmsesilqsgvqse 2526
Qy 1542 HQVYELPDSDHR---LLGDMOK--REDDDPSPYLLAIWTPGETANSAOP-----PEQKCGG 1591
```

```
Db 2527 pa1qplpasqsdct1qasdlr1ressaq1apiaaayds1paandsstspald1apesaqta 2586
Qy 1592 KAS 1594
Db 2587 kat 2589
RESULT 13
ID ABB65772
XX ABB65772 standard; Protein; 5533 AA.
AC ABB65772;
XX 26-MAR-2002 (first entry)
XX Drosophila melanogaster polypeptide SEQ ID NO 24108.
DE Drosophila melanogaster polypeptide SEQ ID NO 24108.
KW Drosophila; developmental biology; cell signalling; insecticide; pharmaceutical.
XX Drosophila melanogaster.
XX WO200171042-A2.
XX 27-SEP-2001.
XX 23-MAR-2001; 2001WO-US09231.
XX 23-MAR-2000; 2000US-191637P.
XX 11-JUL-2000; 2000US-0614150.
XX (PEKE ) PE CORP NY.
XX Venter JC, Adams M, Li PWD, Myers EW;
XX WPI; 2001-656860/75.
XX N-PSDB; ABL09875.
XX New isolated nucleic acid detection reagent for detecting 1000 or more genes from Drosophila and for elucidating cell signalling and cell-cell interactions.
XX Disclosure; SEQ ID NO 24108; 21pp + Sequence Listing; English.
XX The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from Drosophila. The invention is useful in developmental biology and in elucidating cell signalling and cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-ABB72072).
XX The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX Sequence 5533 AA;
SQ
```

```
Query Match 2.4%; Score 221.5; DB 22; Length 5533;
Best Local Similarity 19.1%; Pred. No. 1.4e-06;
Matches 352; Conservative 245; Mismatches 695; Indels 551; Gaps 83;
Qy 5 MDSSAVNATEQNDGS-----RQDVLFEFLNLTPOQKPKRKRKPKVYVVE 53
Db 1362 inns1nasgmgsccsctflpspsr1wrs1shh1qnnh1qg1q1hg1ss1ntclm 1421
Qy 54 GKPKRKRPR-----KPAELPKVVVEGPKRKRKPRKAAATQEKVKSKETGSAKKKLNKESATK 107
Db 1422 aspar-pr1ss1ssnsdsdv1qna1ggs1der1lrnt1een1yer1wsgs1re----- 1471
Qy 108 KPAVNGVMNKSPEVT1KSCRKALNFDLEPNFGDARQGDSESE1VQNSGANSFSE1RDAI 167
```

Db 1472 -----hishtpssatpwsqglmhmlnltglnshqtssagnsnssgt-----vsssa 1520
Qy 168 GGTNGSFLDSVQIDKTNGLGAMNQPLEVSMGNQPKLSTGAKLARQQPDLLTRNOQCQ 227
Db 1521 snsrhkfld-----idel-----qpsdiv-----ksvialkvsfddqfqrlnkqwyd 1563
Qy 228 FPVATONTQPMENQOAWLQKNOLOGFPF-----GNOQPRWTIRNOQPCLA 274
Db 1564 -----psssdalgsnssnltvgsslvavnsrhpqpcsgntspal-----pnla 1607
Qy 275 MGNOQPMYLGTTPRALVSGNOQLGGPOGQNKRPFLNHQTCPLPAGNOLYGSPTDMHQLVM 334
Db 1608 atkatp-iilgn-----csdg--lgnstgksagilqrlssl-----spmspqasm 1650
Qy 335 STGGOQHLLIKNOQPSGLRGOQPCVPLIDQOPATPKGFTHLNOMVATSMSSPGLRPHS 394
Db 1651 sp-----ynspspspvgvgtacilqitkpaaptasagisggtasssp-----a 1697
Qy 395 QSOVPTYLHVESVRILNQTTCQRSRAPAYDSLOQDIHQGNKYITLSHEISNGNCKK 454
Db 1698 ansqptkglgypfshpplntaappavpappplpe---mgkqsrltgg-sggnlntk 1753
Qy 455 AL-----PONS-----SLP-----TPIMAKLEEARGSKROYHRAMGOTEKHDLN 493
Db 1754 slsvpdgpcpsparvqlqksasvpgstnvgapslsldstasvetsasissstsgnss 1813
Qy 494 L-----AQIAQSQDVERHN---SSTCEYLDAAKTKIKQVVQENLHGM-----PPE 538
Db 1814 ltsaahvqkpgqstfveehtksgtstsgssssskkiss-thdklshkhnrrseadk 1872
Qy 539 VIEEDPTDGARKGN--TASISKGAS---KGNSSPVKTAKEKICIVPKTPAKKGRAG 593
Db 1873 kikkdknassdkrknstsgsksatprieddsaadtdak-----aekng-- 1921
Qy 594 RKKSVPFAHASEIQLQOPPPKTPPLSRPKGKGRKSIODSGKARGSGELLCODSIAE 653
Db 1922 -----rhekekerdekreldrkqvereckdrkaqgeerek-----edrkak 1964
Qy 654 IIRYMNQLYGDKEQEQONAMVLYKGDGALVPYVESKKRPRKVDIDDETRINWLLMG 713
Db 1965 -----eeekererekaq-----edrekkereerelekregr-----d 1997
Qy 714 KGDEKDEKDKKKKEKWEERVRFRGRADSFTRMHLVOGDRRFPKSGVVDSVIGV 773
Db 1998 keqkeleirekldrekegrdnrekelr-dkdlrekekrekegr-----ekelhr 2047
Qy 774 FLTQNVSDHLSSAFMSLAFFPKLSSREDERNVSVVDEDEGCILNLNETPSMOEK 833
Db 2048 ekdgrereh-----rekegsrramdveq-egrggrmrelssyqk- 2095
Qy 834 VOHPSDMEVSGVDSG-----SKEQLRD-----CSNSGIERFNFLEK-S 870
Db 2086 ----skndiagaeasltaidqcnkenamtdiagtgpaastpsdntpkersklrns 2141
Qy 871 IONLEEVLSQDSFDPALFQSCGRVSCSCSKSDAEFPPTRCBTKTVSGTSQSVQGTSP 930
Db 2142 pvrlnhrllrsqes-----nhsaggsgcgssghihhedyvkrirmensqnishvsnq 2196
Qy 931 NLSDEICLOGNE-----RPHLYEGS-----GDVQKQET 958
Db 2197 rlndrrdskehkskfkdennssshirphgvcgssassskhhhrdkhhqkgsassiet 2256
Qy 959 TN-----VAQKPKDLEKTMNKNKDSVCFQGNRNDTNWQTPSSSYEQCATROP-HVLD 1009
Db 2257 nssievvvdpisqtkhnl-----ntseeelqshqpkreke 2291
Qy 1010 IEDFGMOGEGLYGSMWISPRVRVKNKNVPRFRFGGVSVPREFTGQIIPSTPHELPGM 1069
Db 2292 rehftshanssssrhks---krdhhhrkhrhsvaestntdehtbpq--qhnphrrisa 2346
Qy 1070 GLSGSSSAVQEHQDDT-----QHNOQDENMKASH-----LQKTFLLDLSSE 1111
Db 2347 agsgageilssaatntsgklhhdhrrsversrsgsgdeghhsskslraklmmllssad 2406

Qy 1112 ECTROSSTKQNTDGLCLPRDRTAEDVVDPLSNSSLQNLILVESNSNKEQTAVEYKETN 1171
Db 2407 sdtddaskkhsifd--ipddcpnvsmgdk-----vkarscknmqraeeekik 2453
Qy 1172 ATILREMKGTLAGCKRPTS-QWDSLRKDVEGNEGRQERNK-----NNMDSIDY 1218
Db 2454 ak-fsqliksgrakkrstsydgs---dte-fedqrhrrsgsssfhgyrpglsssdidd 2508
Qy 1219 EATRASSISISEAIGERGMNMLAV-----RIKOFLEIRIVKHGGIDLEWLRSPDK 1272
Db 2509 eethqrrissdsda-ehggdngdgastladanvrqmqnrlrlcdgdd-----ssede 2561
Qy 1273 AKDYLLSIRGLKSVCEVRLTLHLNLAFFVDTVNGRIAVRMGVPLQPLPESLQLHLLE 1332
Db 2562 irrnmvkhshfgkrnsnstrias-----dse-----sqsqpapd--lciqg 2600
Qy 1333 LYPVLESIQFLWPRICKLDQRTLYELHYLIYFGKVFCTKSRPNCNAC-----P 1382
Db 2601 ehplapa-qeikreqisdeeqk-----fksrhdnsnsseerklktere 2642
Qy 1383 MRGECRHF--ASAYA-SARLALPAPEERSLTSATIPVPPEFPPVAIPMTIELPLPLEKSL 1439
Db 2643 iktelgdgfynsseytytgkikeyspetrkkhks-----krrl 2680
Qy 1440 ASGAPSRENCE-PIIEEPASP-----GOECTETITESIEDIAYNEDDEPTTKLNI 1491
Db 2681 kssstadtsaaqplvmtpltpstpsifdvhsseck--ckfd---nfdlktecsslplel 2734
Qy 1492 EQGEMTLREHMERNMELQE--GDMSKALVALHPTTTSIPTPKLKNISRLRTEHQVVELPD 1549
Db 2735 s--agerikrkerkekkreklrmteatvpnsptndtsekskeer----- 2780
Qy 1550 SHRLDGMKREPDD-----PSPYLLAIWTPGETANSAQPPEQKCGKAS 1594
Db 2781 -hrllkskksmdnscktiynssgahpstpslpa--tptsapstaqlsk-----rge 2832
Qy 1595 KMCWF-----DETSCNSLREANSQTVRGTLIPORTAMRG 1631
Db 2833 dknefifgiisdeeesqfpegaetnkq-----ilpssvsttg 2869
RESULT 14
ABBT1160
ID ABB71160 standard; Protein; 5560 AA.
XX ABB71160;
XX AC ABB71160;
XX DT 26-MAR-2002 (first entry)
XX DE Drosophila melanogaster polypeptide SEQ ID NO 40272.
XX DE Drosophila; developmental biology; cell signalling; insecticide;
XX KW pharmaceutical.
XX OS Drosophila melanogaster.
XX PN WO200171042-A2.
XX PP 27-SEP-2001.
XX PF 23-MAR-2001; 2001WO-US09231.
XX PR 23-MAR-2000; 2000US-191637P.
XX PR 11-JUL-2000; 2000US-0614150.
XX PA (PEKE) PE CORP NY.
XX PI Venter JC, Adams M, Li PWD, Myers EW;
XX DR WPI: 2001-656860/75.
XX DR N-PSDB; ABL15263.

PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
XX interactions -

PS Disclosure; SEQ ID NO 40272; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABU01840-ABU16175), expressed DNA
CC sequences (ABU01840-ABU16175) and the encoded proteins
CC (ABU57737-ABU72072).

CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 5560 AA;

Query_Match 2.4%; Score 221.5; DB 22; Length 5560;
Best Local Similarity 19.1%; Pred No. 1.4e-06;
Matches 352; Conservative 245; Mismatches 695; Indels 551; Gaps 83;

QY 5 MDSSAVNATEATEQNDGS-----RQVLEFDLNTKTPQOKPKRKRKMPKVYVE 53

Db 1362 lnnsslnasqmgscgstflpspsrywssshhqnqnnhqgsgqlhgsssnctlm 1421

QY 54 GKPKRKR-----KPAELPKVVVGKPKRKRKAATQEKVKSKEGSAKKKNLKEATK 107

Db 1422 aspar-prlsnsdsdsvpgnaggsplderlfnfeenywrgsgsre----- 1471

QY 108 KPAVGDMSKSPVTLKSCRKALFNLENPGDARQDSEBIVQNSGANSFSIRDAI 167

Db 1472 -----hisghpssatpswqlsmhmnlstglnshqtasagnsnsgt-----vssa 1520

QY 168 GTNGSFIDSVDQIKTNGLGAMNPLEVSMGNPDKLISTGAKLARDOQPDLLTRNQCC 227

Db 1521 snsrhkfid-----idel-----qpsdiv-----ksvlakksvfdiddfqrlnkqnyd 1563

QY 228 FPVATQNTQFMENQQAWLQMKNLQIGPFF-----GNOQPRMTIRNQPCLA 274

Db 1564 -----pessdalgssnlnvgslnvanvrhpgpcsgntspal-----pnla 1607

QY 275 MGNOQPMYLIGTPRALVSGNQQLGGPOGNKRPFLNHTQICLPAGNQLYGSPTDMHQLV 334

Db 1608 atkatp--lign-----csgg--lgnstgskagllqlrsl-----spmspqasm 1650

QY 335 STGGQHGGLLNQOPGSLIRGOQPCVPLIDQOPATPKGFTHLNQMVATSMSSPGLRPHS 394

Db 1651 sp-----ynspspsvsgvgtacqlgtkpaapgtasaglsqgtaasssp-----a 1697

QY 395 OSQVPTTILHVESVSRILNGTGTQSRAPAYDSLOQDIHQGNKYILSHEISNGCGCK 454

Db 1698 ansqgtklqlyfshpplntcaappavqappipe---mgksritgq--ssgnnltk 1753

QY 455 AL-----PQNS-----SLP-----TPIMAKLEEARGSKRQYHRAMQTEKHDLN 493

Db 1754 slsvdpqpqsparvqlqksasvpgstnvapsslsldsttasvetsasistangns 1813

QY 494 L-----AQQIAQSDVERHN---SSTVEYLDAAKTKYIKVQVQENLHGM-----PPE 538

Db 1814 ltsaiahvqkqpgqstfveehckksgtstsgssssskkiss-thdklshkhnrsesdk 1872

QY 539 VTEIEDDPTDGARKGN--TASISGAS--KGNSSPVKTKAEKEKIVPPTPAKKGRAG 593

Db 1873 kkkksdnassdkrknstsgssksatprtedssaddtdak-----aekng-- 1921

QY 594 RKKSVPFPAHASEIQLWQPTTPKPLRSKPKGKRKSIQDSGKARGPSGELLCCODSTAE 653

Db 1922 -----rhekkerqekredlirkqvereekdrkaqqearek-----edrtrak 1964

QY 654 IITYRMONLYGDKEREQEQONAMVLYKGDGALVPYESKKRPKPKVDIDDETTRINWLLMG 713

Db 1965 -----eeekererekkag-----edrekererelekeqr-----d 1997

QY 714 KDEGEDEDEKDKKKEKWEERRVFRGRADSFIAHMLVQGDGRFSPWKSGSVDSVIGV 773

Db 1998 keqkeirekldrekegrdnrekeir-dkdlrekenrekeqr-----ekelhr 2047

QY 774 FLTONVSHLSSAFMSLAARPPKLSRREDERNVRVSVVEDPEGCILNLFISWQPK 833

Db 2048 ekdqrereh-----rekegrramdveq-egrgrmrrelssyqk- 2085

QY 834 VOHPSDMEVSGVDSG-----SKEQLRD-----CSNSGIERFNFLEK--S 870

Db 2086 ----skmdiaageassltaidcqhnkenamdtiaqtpgaaspstpsdntpkrsrklnrs 2141

QY 871 IONLEEVLSQDSFDPAIFQSCGRVGCSCSKSDAEFPTRCETKTVSGTSGSVOTGSP 930

Db 2142 pvrllhkrllssqes-----nhsaaggsgcgssghqihedyvkkrimensqnlsvhsanq 2196

QY 931 NLSDEICLQGN-----RPHLYEGS-----GDVQKQBT 958

Db 2197 rlndrrdskehsksfkedknssshisrphgcgssasasskhhrrdkhhqkgsassiet 2256

QY 959 TN-----VAQKPDLEKTMNWKDSVCFGQPRNDTNMQTTTPSSSYEQCATRQP-HVLD 1009

Db 2257 nssievvvdpisqtkhnl-----ntseeelqshqkreke 2291

QY 1010 IEDFMQEGELGYSWMSISPRVDRVKNKVPFRFRQGSVPREFTGIIIPSTPHELPGM 1069

Db 2292 rehfsshanssrrhks---krdhhhrkekhrhsvaestntdehtpg-qhnprrisa 2346

QY 1070 GLSGSSSAVOEHQDQT-----QHNOODEMNKASH-----LQKTFDLDLNSEE 1111

Db 2347 agsgagelssaataintsgklhghhrrsverksgsdgghssksklraklmlmsaad 2406

QY 1112 ECLTRQSTKONITDGLPRDRTAEDVVDPLSNSSSLONILVESNSSNKEQPAVEYKTN 1171

Db 2407 sddtdaskkhsifd--ipddcpnvsmyd-----vkaracknmqrqaeekklk 2453

QY 1172 ATILREMGTLADGKKPTS-QWDSLRLKQVGEQNEGRERK-----NNWDSIDY 1218

Db 2454 ak-fsqliqsrakkrstsydgds---dte-fedqhrnsgsssfhrypglssddddd 2508

QY 1219 EAIRRASISEISEAIKERGMNNMLAV-----RIKDFERIVKDHGIDLEWLRSPDPK 1272

Db 2509 eethhrrissdsda-ehaggdnggastladanrvqmqnlrrlcdgd-----ssede 2561

QY 1273 AKDYLLSIRGLGKSVCEVRLTLNLNLAFFVDTNVGRIAVRMGWVPLQPLPSQLHLE 1332

Db 2562 lrrnmkshfgrksnstrias-----dse-----sgsqgpad---ltikq 2600

QY 1333 LYPVLESTOKFLWPRLCKLDQRTLYELHYQLITFCGVCTKSRPNCAC-----P 1382

Db 2601 ehpiapa-geikreqlsdeeq-----fksrhdnsnsieerklktere 2642

QY 1383 MRGEGRHF--ASAYA-SARLALPAPEERSLTSATIPVPPESPVAIPMIELPLEKSL 1439

Db 2643 lktelgdlynsseytygtklkeysptekkhks-----krll 2680

QY 1440 ASGAPSNRENC-PPIIEBPASP-----GOECTEITSDIEDAYYNEPDPDPIPTKINI 1491

Db 2681 ksstadsaaqtplvmtpltpsfidvhsseck--tkfd-----nfdldlkecsaplei 2734

QY 1492 EQFGMTLREHMERNNLEOE--GDMSKALVALHPTTISPTPKLNISRLRLEHVOYELPD 1549

Db 2735 s--agerrkhrkekkrklrnmteatvpnsptndtsseklskeer----- 2780

QY 1550 SHRLDGMKKRPDD-----PSPYLLAIWTGCTANSQAQPPQKCGGKAS 1594

Db 2781 -hrkkkskksmdnsctnklynssgahpstpsipa--tptsapstaqtsk-----rge 2832

QY 1595 GKMF-----DETSECNSLREANSQTVRGTLTLLIPCRTAMRG 1631

Db 2833 dkmeffilgsdeesqfpeaetnkd-----lipssvsttg 2869

RESULT 15

AAU14603
ID AAU14603 standard; Protein: 5373 AA.

XX AAU14603;

DT 24-OCT-2001 (first entry)

XX Novel bone marrow polypeptide #2.

XX Bone marrow; diagnostic; therapeutic; gene therapy; antigenic;
XX haematopoiesis; myeloid; lymph cell disorder; tissue regeneration;
XX wound healing; nutritional supplement; immune disorder;
XX severe combined immunodeficiency; SCID.

OS Homo sapiens.

XX WO200157187-A2.

XX 09-AUG-2001.

XX 05-FEB-2001; 2001WO-US03782.

XX 03-FEB-2000; 2000US-0496914.

XX 20-JUN-2000; 2000US-0598075.

XX 19-JUL-2000; 2000US-0620325.

XX 30-NOV-2000; 2000US-0250683.

XX (HYSE-) HYSEQ INC.

XX Ford JE, Boyle BJ, Tang YT, Liu C, Asundi V, Zhou P, Xue AJ;

PI Ren F, Drmanac RT;

XX WPI: 2001-488875/53.

DR N-PSDB: AAS22908.

XX Nucleic acids encoding bone marrow polypeptides, useful in diagnostic

PT and gene therapy -

PT Claim 10; Page 198-208; 392pp; English.

PS AAU14602-AAU14794 represent novel bone marrow polypeptides of the

XX invention. The proteins and corresponding coding sequences may be used

CC in the prevention, diagnosis and treatment of diseases associated with

CC inappropriate bone marrow polypeptide expression. For example, to treat

CC disorders associated with decreased expression by rectifying mutations

CC or deletions in a patient's genome that affect the activity of the

CC polypeptides by expressing inactive proteins or to supplement the

CC patient's own production of the polypeptide. Additionally, the nucleic

CC acids may be used to produce the polypeptides, by inserting the nucleic

CC acids into a host cell and culturing the cell to express the protein.

CC The nucleic acid and its complementary sequences may also be used as DNA

CC probes in diagnostic assays to detect and quantitate the presence of

CC similar nucleic acid sequences in samples, and therefore which patients

CC may be in need of restorative therapy. The proteins may also be used as

Query Match 2.4%; Score 216; DB 22; Length 5373;

Best Local Similarity 17.5%; Pred. No. 3.6e-66;

Matches 322; Conservative 292; Mismatches 698; Indels 524; Gaps 77;

Qy 16 TEQNDGSRQDVLEFDLNKTPQOKPSKRKRKFKMPKV-----VVEGKPKRKRKPAELPKV 69

Db 1542 ttqddlsalqkngsdll-kdlqddiqnratstfatvvkdiegmeenqtklspreltal--- 1597

Qy 70 VVEGKPKRKRKAATQBKVKSKETGSAAKKNNLKESATKKPANVGDMGNKSPSEVTLKSCRK 129

Db 1598 -----reklhakeqyealqeetrvaqke-leeavt-----s 1628

Qy 130 ALNFDLENPGDARQDSESEIVQNSGANSFSEIRDAIGGTNGSFLDSVSOLDKTNGLGA 189

Db 1629 alqgeteksaak-----elaenkkidalldwvtvsgsggqlllnlpgmeglsasl 1682

Qy 190 MNOPLEYSMG---NQ-PDKLSTGAKLARDOOPDLLTRNQOCQFPVATQNTQFFMENQQA 244

Db 1683 ekgalddtdgymgvnqapeklldkqcemmkarhqell--sqgnfilatqsaqafldqghg 1740

Qy 245 WL-----QMKNNQLIGFPFGNQOPRMTIRNQOCPCLAMNQOPMYLIGTPRPALVS----- 293

Db 1741 nltpeeqmllqklg-----elkegystslagseaealkqvltqdelqkflqdhk 1790

Qy 294 -----GNQQLGGPQGNKRPIFLNHQTCLPAGNQLYGSPTDMHQLVMSTGGQQ 340

Db 1791 efeswlersekelennhkggsspetlpsllkrq-----gs-----fsedvishkgdl 1837

Qy 341 HGLLIKNOQ---PGSLIRGOQPCV---PLIDQOPATPKGFT-----HLNQMVA 382

Db 1838 rfvtsigqkvldmensfkegkepsignlvkdkldaterytalhskctrllgshlnmlig 1897

Qy 383 TSMSSPGLRPHRSQSVPTTYLHVESVRILNLTGTCQSRAPAYDSLQDDIHQNKYIL 442

Db 1898 -----qyhqfqnssadsldqwmqacean-----veklstsvsdpgvl 1935

Qy 443 SHEISNGNGCKKALPONSSLPPTIMAK-----LEEARSKROYHRAMGOTKHDNLAAQOI 498

Db 1936 qeqlattkqlqeelaeh-qvpveklqkvardimelegapdhvqgettdsilshfql 1994

Qy 499 AOSQDVERHNSSTCEVYLDAAKTKIOKV-----VQBNLHGMPEVEIEIDDDTDGARK 552

Db 1995 s-----yslaerssllqlaagsvqseslesllqsgiveeqn-----le 2034

Qy 553 GKNTASTISKGASKGNSPVKKTAEKECIVPKTPAKKGRAGRKSKSVPPPAHASEIQLNQP 612

Db 2035 gkvyslssgvi-----qkalatnmklqdiarkksl--eatremvtrfme 2079

Qy 613 TPPKTPILSRSPKGRKSIQDSGKARCPSELCCQDSIABLIYRMQNLVILGDKREREQ 672

Db 2080 tadattaa-----vlgklaevsqrfeqlclqgqek----- 2110

Qy 673 NAMVLYKGDGALVPVEKKRPRKVDIDDETT-RIMNLLMGKDEKDEKDKKKEKW 731

Db 2111 -----eslklqlpqaemfehlsgklqgfmenksrmlasngpqddihf 2155

Qy 732 WEERRRVRGRADSFIAARMHLVQGDRRFSPWKSGVSDVIGVFLTNQVSDHLSASFNSL 791

Db 2156 fqq-----iqelnlemedqgen-----ldtl-----ehlvitelsscgfald 2191

Qy 792 AARPPPKLSSSRREDNRNVRVVE---DPEGCILNLE-----IPSQOEKVQ---HPS 838

Db 2192 lcqhqrqvnrlkdfteqlqtkvkerekdasscqeqelrkvlrtfqtwlketeqsppt 2251

Qy 839 DMEVSGVD-----SGSKQLRDCNSG---IERFNFLEKSIONLEEVLVS---SQDSFDPAI 889

Db 2252 etsvsakeleqiehlkldldwaskgtlveeincnkgtslenlmeitapdsqgktdlte 2311

Qy 890 FQSCGRVSGSCSCSDAEFF-----TTRCET--KTVSGTSGSVQF-- 927

Db 2312 lq-----cdmsdvnlkyeklvgvlhergeslqalnlrmeevhkeansvlqleske 2362

Qy 928 -----GSPNLSDEICIQGNER-----PHLYEGSGDVQ--KQETTNVAKKPKDLEXTM 972


```
Db 2363 evlksmdamsptkktetvkaqaesnkafaelaeqnsqkikvkealagllvtypnsgeae 2422
QY 973 NMKDSVCFQPRNDNW-----QTPSSSYEQ-----ATROPHVLDIE-DFG 1014
Db 2423 nwkkl-----qeelnrweratevtvarqqlaesashlacfaaesqlrplwimekelmng 2478
QY 1015 MOGEGLGYSWMSISPRVDRVKNKNVPRFRFGGVSVPREFTQIIIPSTPHELPGMG-LSG 1073
Db 2479 vlgp-----lsldpumlnaqkqv--qfmkfearrqgheqlneaagqiltpgdvs1 2530
QY 1074 SSSAVQEHODDTQHNOQDEMNKASHLOKTFDLNLSSECLTRQSTKONITDGC1PRDR 1133
Db 2531 stsqvqke-----lqsinqkwvcltdklnsrssqldqalvkt 2568
QY 1134 TAEDVVVDPLSNSSL--QNILVESNNSKEQTAVEYKETAATILREMGKTLADGRKPTSO 1191
Db 2569 qyqellqlisekvragqrlsvqsaistqpeavkqqlseetseirsdieqlchevkeaqtl 2628
QY 1192 WDSLRKDVEGNEGROERNKNNMDSIDYEAIRRASISEISEAIKERGMNMLAVRIKDFLE 1251
Db 2629 cde1svli-geqylkelkkrietv-----alplggledlaadrinriqa 2672
QY 1252 RIVKDHGGIDL-----EWLRESPPDKADYLLSIRGLGLKS-----VECVRLTLHLNLA 1301
Db 2673 alastqfqgmfdelrtwldkqsqakncplsa1erlqsqlgeneeefqkslnqhsqsy 2732
QY 1302 PVDTVNVIARVMGWVLOPLPESQLHLLELYPVLESIQKFWPRLCKL----DQRTLY 1357
Db 2733 evlvaegesallls--vppgeekrtlqnqlvelknhweelskktadrqsrkdcmkakqy 2790
QY 1358 ELHYQLITFGKVFCYKSRPNACPMRGECRHFASAYASARLALPAPERSTSATIPVP 1417
Db 2791 qwhve-----dlvpwiedck---akmselrvtld----- 2816
QY 1418 PESFPVVAIPMIELPLLEKS-LAGAPSN---RENCEPIIEEPASPGQECTEITESDI 1472
Db 2817 -----pvqlessllrskamlnevekrsllelinsadlinsseadedgi 2862
QY 1473 ED--AYYNEDPD-----EIPTIKLNIEQFGMTLREHME--RNMELQ-EGDMSKALV--ALH 1521
Db 2863 rdekaginqmdavteeelqaktsleemtqrlrefgesfkniekkvegakhqlleifdalq 2922
QY 1522 PTTTSIPTPKLNISRLRTEHQVYELPDS-----HRLDGMKREPD--DPSPYLLAIWT 1574
Db 2923 sqacs-----nknlekiragqevlqalepqvdylnftqglvedapdgdsasqll----- 2972
QY 1575 PGETANSAQPPEQKCGKAGKMKCFDETCSECNLSREANSQTVRGTLIPCRRTAMRGSP 1634
Db 2973 -----hqaevaqge-----flevkqrwns-----gcvmmenkleigiqf- 3006
QY 1635 LAGTYFQVNWELFADHESSLKPIDVPRDWIWDLPRTT 1670
Db 3007 ----hcrvremf----sq1adlddelgmggaigrdt 3034
```

Search completed: September 27, 2002, 12:54:41
Job time: 10230 sec

THIS PAGE BLANK (USPTO)

; TELEX: 25-3856

; INFORMATION FOR SEQ ID NO: 5:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1780 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

us-08-769-309A-5

Query Match 2.1%; Score 195; DB 1; Length 1780;

Best Local Similarity 17.8%; Pred. No. 7.6e-07;

Matches 332; Conservative 262; Mismatches 690; Indels 584; Gaps 84;

Qy 105 ATKPNYGMNSKSPVTLKSKRKALNFDLENPDARQDSSEIVQNS-----SGAN 158

Dy 23 AEPEPGGGPSAEAPDPT-----ADPAIA-ASDPATKLLQKNGQLSTINGVA 69

Qy 159 SFSEIRDAIGTNGSFLDSQIDKTNGLGAMNQPLE-----VSMGNQPD 203

Dy 70 EDELSLQEGDLNG-----QKALNGQAGALNSQEEIEVTEVQRDSEDSERDSK 122

Qy 204 KLSTGAKLARQDQDOLLTRNQOC-QFFVATON---TOFPMENOQAWLOMKN--QLIGFP 256

Dy 123 EMATKSAVVHIDTDCQENRNIEQIPSESNLELTQ-PTESQANDIGFKVKFVGFK 181

Qy 257 FGNQOPR-----MTIR-----NOQPCILAMG-----NOQPMYLIGTPRPA 290

Dy 182 FTVKKDKTEKPDVTQLLVKKDEGEAGAGDHQDPSLGAAGEAASKSEPKQ--STERKE 239

Qy 291 LVSGNQQLGGQGNKRPIFLNHQTLCPAGNQLYGSPTDMHQLVMTGGQQHLLIKNQOP 350

Dy 240 EPLKREQ-----SHAETSP-----PAESGQAVERC--KEEGEEKQKEPE 276

Qy 351 GSLIRGQOPVPLDQOPATPKGFTHLMQVATSMSSPGLRPHSOSQVPTTYLHVESYR 410

Dy 277 SK--SAESPTSPVSETGSTFKFP--FTQGWAGWRKKTSPFRKPEDEVAS----- 323

Qy 411 ILNGTTGTGCRSAPAYDSLOQIHQGNKYILSHEISNGCKKALPONSSLUPTPIAKL 470

Dy 324 -----EKKKEQEKEKVDTEED--GKAEVASEKLJTASEQAHPQEPASAEHPRLSAEY 373

Qy 471 BEAR-GSKROYHRANGQTEKHDNLNLAQOIAQSODVERINSS-----TCVEYLDAAKT 522

Dy 374 EKVELPSEEQVSGSGPSEKPAPLATEVF-DEKIEVHQEEVAVVHVSTVEERTEEQK 432

Qy 523 KIQK-----VVOENLHGMPEVEIEB-----DQPTDGARKGNTASISKA 563

Dy 433 EVEETAGSVPAEELVGMDAEPQAEPAKELYKLKTCVSGEDPTQGADLSPDEKVLSP 491

Qy 564 SKGNSSPVKKTAEKCIIVPKTPAK-----KGRAGRKKSVPVPPAHASEIQLMQPTPP 615

Dy 492 PEGVYSEVEMLSQBERMKVQGSPLKLTSTGLKLSCKKQKGRGGGDEESGEHTQVPA 551

Qy 616 KTLPLSRPKGKGRKSQDSKARGPSELIC-QDSIAEIIYRMQNLVGLGKERQEQNA 674

Dy 552 DSPDSQEQKQKGESSASPEEPE-----EITCLEKGLAEV-----QQDGEAEEGA 595

Qy 675 MVLYKGDG-----ALVPYES-----KKRPRP-KVIDDITTRIMNLNG----- 713

Dy 596 -----TSDEKKEKREGVTPWASFKKMWTPKKVRRPSESKDELDKVSATLSSTESTASE 651

Qy 714 -----KGDEKEGDEBKKKK-----EKWWE-----BEER 736

Dy 652 MQEEMKGSVEEPKPEPKRKVDTSVWEALICVGSSSKKRRRRSSSDEEGPKAMGGDHQ 711

Qy 737 RVFRGRASFTARMILVOGDRRFPKWSGVDSVTGVLFTQNVSDHLSS-SAFMSLAARF 795

Dy 712 KADEAGKDKETGTGDIGLAGSQEHDPGQSSSPEQAG-----SPTGEGVSTWESFKRLVT-- 766

Qy 796 PPKLSSSRREDERNVRSV----- 813

Dy 767 PRKSKSLEEKSEDSIAGSGVEHSTPDTEPGKEESVWSIKKFTIPGRKKRPDQKQEQAP 826

RESULT 3

US-08-994-570-5

; Sequence 5, Application US/08994570

; Patent No. 6090929

APPLICANT: Su, Xin-zhaun
APPLICANT: Wellens, Thomas E.
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESS:
ADDRESSEE: Knobe Martens Olson & Bear
STREET: 620 Newport Center Drive 16th Floor
CITY: Newport Beach
STATE: California
COUNTRY: US
ZIP: 92660
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/568,459A
FILING DATE: 07-DEC-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Israelsen, Ned
REGISTRATION NUMBER: 29,655
REFERENCE/DOCKET NUMBER: NIH121.001CP1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1435 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ORGANISM: Plasmodium falciparum
US-08-568-459A-4

Query Match 2.1%; Score 188.5; DB 2; Length 1435;

Best Local Similarity 18.3%; Pred. No. 1.9e-06;
Matches 185; Conservative 144; Mismatches 345; Indels 337; Gaps 42;

QY 441 ILSHSINGNGCKALPQNSLPTPIAK-----LEEAGSKRQYHRAMGQTEKHDLNL 494
DB 299 MLSEHKNNINCKN-IPQELQITQIKWHGFEFLERDRNSKLPKSKKNNT----- 350
QY 495 AQTAQSQDYVRHNSSTCVYLD-----AAKTKIQKVVQENLHGMPPVEIEIE 543
DB 351 -----LYEACEKECIDPKMYRWILRSKFHWHTLSKEYETQKVPKNAENY---LKIS 402
QY 544 DDPDYGARK-----GKNTASIGKASKGNSPVKKTAKEKCIIVKTPAK 588
DB 403 ENKNDKAVSLLNNDCAEYSKYCDCKHTTLVKSVLNGNDNTIK---EKREHIDLDDFSK 459
QY 589 KGRAGRKSVPPAHASEIOLQWPTPKTPLSRKPKGKRKSTQDSKGARGSGELLQ 648
DB 460 FG--CDKNSVD-----TNTKWCKNPYI-----LSTKD-----VC- 488
QY 649 DSAEIIYRMQNLGLDKREQEQNAMVLYKGDALVPYESK--KRKPRPKVD-----ID 701
DB 489 -----VPPRQELCLGNIDRIYDKNLLMIKEHILAIYESRIILKRYKKNDDKEVKII 543
QY 702 DET-----TRIWN-----LLMGKDEKGEDEKDKKKEK-----WVEERR----- 737
DB 544 NKTADIRIDIGTDYVNDLSNRKLVGKINTNSKYVHRNKNKDLFRDEWVKVKKDVWN 603
QY 738 -----VFRGRADSFARMHLVQGRDRRSPKGSVVDSYIGVFLTQNVSDHLSLSAFAWSLAA 793
DB 604 VISWVFK-----DKTVCKEDDIENIPQFRWFSWGGDY----- 637

QY 794 RFPKILSSSRREDERNVRSVVVE-----DPEGCILNLINEIPSWOEKVQHPSPDMEVSGVDS 847
DB 638 -----CODKTKMETLKVECKEPCEDNDCKNCKSCYKWKISKKK----- 677
QY 848 GSKEQLRDCSNSGIERFNFLEKSIQNLLE-----EVLSSQDSFDPATF-----QSCGRVGS 898
DB 678 -----EYNKOAKQYQYQKGNKMYSEFKSIKPEVLYLKYSEKCSNLF 723
QY 899 CSCSK-----SDAEFPTRC-ETKTV-----SGTSQSVQSGPNLSDEICLQGNRPHLYE 948
DB 724 EDEFKEELHSDYKNCCTMCPEDVDPVLSIIRNNEQTSQEAVPPEENTEIA-HRTETFSISE 782
QY 949 G-SDYQKQ-----ETTINVAQKKPDLEKTMKMKDSCVCFQPR 984
DB 783 GPKGNEQKERDDSLSKISVSPENSRPETDAKDTNLLKLKGDVDISM---PKAVIGSSP 839
QY 985 NDTNWQTTSSSYEQCATR-----QPHVLDIEDFGM-QGEGLYGSHMSISPRV----- 1031
DB 840 NDNINVTQGDNISGVNSKPLSDVDPKKELEDQNSDESEETVYVNHISKSPSINNGDSS 899
QY 1032 -----DRVKNKNVPRFRFROGGSVPRE 1053
DB 900 GSGSNAVSESSSNTGLSIDDRNGDTFVRTQDTANTEDVIRKENADKDEKGADEERH 959
QY 1054 FTQIIIPSTPHELPGMLSGSS---SAVQEHQDDTQHNQOD---EMNKASHLQKTF--- 1103
DB 960 STSESLSPEEKMLTDNEGNSLNHEEVKEHTSNSDNVQSGGIVANNVKEKLDLTENP 1019
QY 1104 ---LDLLNSEECLTRQSTKQNTGCLPRDRTAEDVDPLSNSSLONLIVESNSSNK 1160
DB 1020 SSSLDEGKAHEELSEPNSDDQMSNTGPLDNTSEETTERISNNEYKVNREDERTLTK 1079
QY 1161 EGTAVEYKETNATILREM-KGTIADGKKPTSQWDSLRKQVE---GNEGRQ----- 1206
DB 1080 EYEDIVLK---SHMNRSDDGELYDENSLLSTVNDSEDEAEAMKMGNDTSEMHSNHSQHI 1136
QY 1207 --ERNKNMDSID-----YEAIRRASISEISEAIAKER 1236
DB 1137 ESDQKNDMKTVGDLGTHVQNEISVPVTGEIDEKLRKESKESKIHKAERER 1187

RESULT 6

US-08-487-826B-4

; Sequence 4, Application US/08487826B

; Patent No. 5993827

; GENERAL INFORMATION:

; APPLICANT: Sim, Kim L.

; APPLICANT: Chitnis, Chetan

; APPLICANT: Miller, Louis H.

; APPLICANT: Peterson, David S.

; APPLICANT: Su, Xin-zhaun

; APPLICANT: Wellens, Thomas E.

; TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX

; AND PLASMODIUM FALCIPARUM ERYTHROCYTE

; BINDING PROTEINS

; NUMBER OF SEQUENCES: 45

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Knobe Martens Olson & Bear

; STREET: 620 Newport Center Drive 16th Floor

; CITY: Newport Beach

; STATE: California

; COUNTRY: US

; ZIP: 92660

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent In Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/487,826B

; FILING DATE: 10-SEP-1993

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Israelsen, Ned

Db 63 LMKREMCVSGDHTQTLTDLSPSEKTLKPHKEGIVSEVMSLSSQERIKVQSGPLKLFSS 122
Qy 388 PGLRPHS-----QSQVPTTYLHVSVSRILNLTGTCQSRAPAYDSLQOQIHQ 436
Db 123 SGLKKLGGKQKRGKGDEPEYQHITHES-----PESADEQK 163
Qy 437 GNKYILSHEISNGCKKALPQNSLPTIPMAKL--BEARGSKROYHRAMGOTEKHDNLN 494
Db 164 GESSASSPEEPETCLEKGP---LEAPRMGLURKELLRGEKKR-----KDHSLGI 211
Qy 495 AQAQASQDVERHNSSTCVELYDAAKTKTKQVVQENLHGMPPVIEIDD-----545
Db 212 LOKDGTQETVRRPS-----ESDKEELEKVKASATLSLSDTSTVSEMDEVKTVGBEQK 264
Qy 546 PTGARKGNFTAS---ISKASK-----GNSSPVKKTAKEKC---580
Db 265 PEPKRRVDTVSNEALICVSSKKRARKASSDIRPRTILGGQOSRGGQQORSRD 324
Qy 581 IVPKTPAKKGRGRKKSVPP---PAHASEIQLWQP-----TPPKTPLSRKPKGKGRKSI 632
Db 325 AVPASTQEQDAQGSSSPAGSPSEGEVSTWESFKRLVTPRKKSCKLEKEAGRTL 384
Qy 633 QDSG-----KARPSGELLICQDSIAETIYRMQNLILGDK-EREQEQNAM 675
Db 385 VGACPLRSNRVEKNLGFPLRNSPDG-----GRKGOMGRQSOAT 424
Qy 676 VLYKGDGALVPYSEKRRKPRKVDIDDETRINLLMGKDEKDEKDKKKEKMWEE 735
Db 425 VEDSG-----DVPVAVPL-----SEYDAVEREKME-----458
Qy 736 RVFRGRADSFIAHMLVQGRDRFSPWKGVSDVIGVFLTONVSDHL--SSSAFMSLAA 793
Db 459 -----AQGNALPSCWGVVSELSKTLVHTVSVAVIDGTRAVTSVEE 501
Qy 794 RPPKLSSS-----REDERNRSVVVED-----PECCIILNLNIPSWQ 831
Db 502 RSPASISASVTEPLEHTAGEAMPVEEVTEKDIITAEETPVLTQLPEG-----549
Qy 832 EKVQHPSDMEYSGVD-----SGSKE-----QLRDCSNS 859
Db 550 -KDAH-DMVTISEVDFTEAVTATSETSEALRTEEVTEASGAETITDMVSAYSQLTDSPT 607
Qy 860 GIER--FNFLEKSTONLEEVLSQDSFDPALFQSCGRVSCSCSKSDAEPTTRCETKT 917
Db 608 TEATPVOQVEGGVLDTEEROTQ-----AILQAV-----ADKVKESQVPATQTVORT 657
Qy 918 VSGTSQVQSGPNLSDEICLQGNRPHLYEGSGDVQKQETTNVAOKPDLKTMNRKDS 977
Db 658 GSKALEKVE-----EVEDESEVLASEKEKDVMPKGPVQAGAEHLAQ-----699
Qy 978 VCFGQPRNDTNQTTSS-----SYEQCATRPHVLDIEDFGMGQEGGLGYSMNSTSP 1029
Db 700 -----GSETQATPESLEVPENTADVHVATCO--VIKILQOLMEQ-----AVAP 741
Qy 1030 RVDVKNKNVPRFRQGSVPREFTGIIIPSTPHELFCMGLSGSSSAVQEHQDTHQNO 1089
Db 742 ESSETLTD-----ETNGSTP-----LADSDTADGTQQDETIDSO 776
Qy 1090 QDMNKASHLOKTFDLNLSSECLTROSSTKQNTIDGCLPRDRAE-----DVEDPLSN 1145
Db 777 DSKATAAARQSOV-----TEEAATAKEEPTSLPNNVPAQEEHGEPPGRDVLPTQOE 830
Qy 1146 SSLQNILVESNSNKEQTAVEYKETNATILRMKGTLAGKK-----PTSQ--1191
Db 831 LAAAAPVWQKTEVQGEQEDV-----LDGEKVKEEQEVFVHSGPNSQKA 875
Qy 1192 -----WDSLURKQVEGNEGQERKNMWSIDYEA-----IRRASISESEALKERGMNM 1241
Db 876 ADVTYDS---EYMGVAGQCEKSTEVQSLSEEGEMETDVEKRETRPEQVSEEGEQT 932
Qy 1242 LAVRIKDFLERIVKHGGIDLEWLRESPPDKAKDYLLSTIRGLKSVCEVRLTLTHNL--1299
Db 933 AA-----PEHERNYGKPVLTLDMPSS--RCKALGSLGG 964

Qy 1300 --APPVDTNVGRIAVRMGWVPLQPLPESLQHLLELYPVLESQKFLWPRCLKLDQRTLY 1357
Db 965 SPSLPQDKAGCIEVQV-----QSLDTTWTQ--TAEAVEKVI-----999
Qy 1358 ELHYQLITFGKVFCTKSRPNCNACPMRGECHFPASAYASARLALPAPEERSLTS--1412
Db 1000 -----ETVVIS--TGESPEC-----VGAHL-LPA-EKSSATGGHWTLO 1034
Qy 1413 ---TIPVPESPFPVAIPMIEPLPLEKSL-----ASGAPSNRENCEPIIEEPASP---1460
Db 1035 HAEDTVPLGPES-QAESIPILVTPAP-ESTLHPDLOGEISASQORSEEEKKPDAGPDAD 1092
Qy 1461 GOECTEI-----TESDIEDAYNDDPDEIPTIKLINIQFGMT-----1497
Db 1093 GKSTADIKVLKAEPEILELESKNKIVLNVIQTAVDQFARTETAPETHAYDSQTVQVAM 1152
Qy 1498 -----LREHMERNMELOEGDMSKALVALHPTTTSI-----1527
Db 1153 RLDSRENRWCWTKMKVAKMKHPVQPPREDIQVLTVLEAWLSSEMLAALAVESAGVKYSIE 1212
Qy 1528 ---PTPKLKINISRLRTEH-----QVYELPDSHRLLDGMDKREPDDPSPYLLAIWTPGETA 1579
Db 1213 KLPPQPKDQK-----EHAADGPQLQSLAQAEAVSGNLTKESPDTNGPKL-----1256
Qy 1580 NSAPPEQKCGGK--ASGKCMCFDETCSE-----CNSLRE-----ANSQTVRGT 1620
Db 1257 ---TEERCPCQLRSRKKKCLPSQSRTRPRQKRTCSRQRETWQNPKNMLVAHCTSRVP-1310
Qy 1621 LLIPCRTAGMSFPPLNGTYFQVNFELFADHE 1650
Db 1311 ---ECENKSONKMILLGPWTKISEPMKRSRE 1337
RESULT 8
5180808-2
; Patient No. 5180808
; APPLICANT: RUOSLAHTI, ERKKI I.
; TITLE OF INVENTION: VERSION CORE PROTEIN, NUCLEIC ACID
; SEQUENCES ENCODING THE SAME, NUCLEIC ACID PROBES, ANTI-VERSION
; ANTIBODIES, AND METHODS OF DETECTING THE SAME
; NUMBER OF SEQUENCES: 4
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/441,179
; FILING DATE: 27-NOV-1989
; SEQ ID NO:2
; LENGTH: 2409
5180808-2
Query Match 2.08; Score 185.5; DB 6; Length 2409;
Best Local Similarity 17.8%; Pred. No. 8.2e-06;
Matches 315; Conservative 234; Mismatches 624; Indels 597; Gaps 81;
Qy 103 ESATKPKANVGDMSNKSPEVTLKSCRKALNFDLENP--GDARQGDSESEIVQNSGANSF 160
Db 403 EEEEECANADVT--TPSQVQYINGKHLVTPKDPPEAEARRGQFESVAFSONFSDSE 461
Qy 161 SEIRDAIGTNGSFLDSVSDIKTNGLGMNQ-----LEVS-----M 198
Db 462 SDTHPFV-----IAKTE-LSTAVQPNNESTETTESLEVTKPETYPTSEHFS 507
Qy 199 GNQDPKLT-----GAKLARDQOPDLLTRNQOQFPVATQNTQFPHENQ-Q 243
Db 508 GGEQVFVTPVTFHEEFESGTAKKGAESVTERDTEYGHQAEHTEPVSL-----FPEESGE 563
Qy 244 AWLQMKNLQIGF-----PFGNQOPRMTIRNQOCPCLA-----MGNOQPMVLTGTPRPA 290
Db 564 IADQESQKIAFARATEVTFGEEVEKSTSVITYPTIVPSSASAYSEEAATLIGNPWP 623
Qy 291 LVSGNQQLGGQPGQKRRPFIPLAHQTCPLA-----GNOLYGSPTDMHOLVMTSGQQ 340
Db 624 DLLSTKESWVEATPRQVVVELSGSSSIPITEGSGEAEDEDTMTMTVTDLSQ-----674


```
QY 197 --SMGNOPDKLSTGAKLARDQOQDPLLRNOCQOPVATONTQPMENQAWLQMKNLIG 254
Db 520 LNALGNPMSPVAGG-ITTOQPPNLSALPTSIGATN---PLMD-----G 564
QY 255 PFPGNOQPRMTIRNOQCLAMGNOQPMY-----LIGTPRAL----- 291
Db 565 SNSGNIGSLSTIPTAAPSSSTGVKRGHWHVTQDLRHLVHKLVAQAFPTDPDAALKDRR 624
QY 292 -----VSGNOQLG 299
Db 625 MENLVAYAKKVEGMYESANSRDYHLLAEKYYIKQKELEKRRRLHLKOGILGNOPAL 684
QY 300 GPQCNKRPFIPLNHQTCPLPAGNQLYGSPTDMHQLVMSTGGQGHLLIKNOQPGSLIRGOQP 359
Db 685 PASGAQPPVPPAGSVRPPNCL---PLPNRMQVSG-----MNSFPMSLGNVQLP 734
QY 360 CVPLIDQOPATPKGFTHLNOMV-ATSMSS-PG--LRPHSOSQVPTTYLHVSVSRILNGT 415
Db 735 -----QAPMGPRASPNNHVSQMNMSASVPGMAISPSRMPQPPN-----MMGT 777
QY 416 TGTQCRSAPAYDSLOQDIHOGNKYILSHSLSNGCKKALPQNSSLPTPIMAKLEEAR 475
Db 778 HANNIMAQAPTONQF-----LPQN----- 796
QY 476 SKROYHRAMGQTEKHDLNLAQIAQSODVERHNSSTCVE-----YLDAAKTKRIQKVQE 530
Db 797 --QFPSSGAMSVNSGMPGPAAGVSGQEPGAALPNPLNLAPOASQLPCPPVQTQS 853
QY 531 NLHGMPEVIEIDDP-----TDGARKGNTASISKASK 565
Db 854 PLHPTPPASTAAGMPSLQHPTAFGTPPQAPATQPTSTPVSSGOTPTTPGSPSAQOT 913
QY 566 GNSPVKTAKEKCIKPTAKKGRAGRKSVPPAHASEIQLWQPT-----PPKTPLS 620
Db 914 QSTPTVQAAAQVTPQOTPVQ-----PPSVATPOSSQOQTPVHTQPPGTPLS 963
QY 621 RSK-----PKGGRKSIQDSKGARPGSGELLCODSIAEIIYRMQNLVIGDKEREQON 673
Db 964 QAAASIDNRVPTSTVISAETSSQOPGPDVPL-----EMKTEVQT---DDAEPE----- 1010
QY 674 AMVLYKGDALVPYESKRRKRPVKVDDIDETRIWNLMLGKDEKDEKDKKKEKWE 733
Db 1011 -----PTES-KGEPRSEMMEED-----LQSSQVKEETDTTEQKSEPEV 1049
QY 734 BERRVERGRADSFITARMHLVQDRRRFPWKGVSDSVIGFELTONVSDHLSSSAFMSLAA 793
Db 1050 EEK-----PEVKVEAKEEENSSNDTASQSTSP 1078
QY 794 RFPPLKSSRDER-----NVRSVVDEPGCIL-----NLNEIPSWQEKVQHPSDMEV 842
Db 1079 SOPRKKIEKPEELQALMPTLEALYRQDPESLPFRQPDQLLIPDYFDIVKPNMDLST 1138
QY 843 --SGVDSKSEQ-----LRCSNSGIERFNFLEKSIQNLLEEVLSSQDS 884
Db 1139 IKRKLDTOYQEPQWYVDVRLMFNNALYRKTSRV--YKFCSKLAEVFEQEI----- 1190
QY 885 FDPALFQCGRGVSCSCSKSDAEPTTRCETKTVSGTSGVQTPGPNLSDSIC-----L 938
Db 1191 --DP-VQOS---LGYCCGRKYEFSPOTLCYCKQKICTIPRDAAYYSQNYRHFCCGKCFEI 1245
QY 939 QGNERPHLYEGSDGVOKETTNAVQKPDLEKTMN----- 973
Db 1246 QG-ENVTL---GDDPSQPTTIS---KDOFEKKKNDTLDPEPFVDCCKGRMKHQCIVLH 1298
QY 974 -----WKDS-VC-----FGQPRNDTNW-----QITPSSSY-----EQCATQPHVDLIED 1012
Db 1299 YDIWPSPGVCDNCLKTKGRPKENKFSAKRLQTTLRGLNHLEDVRNKFLLRQNHPEAGEV 1358
QY 1013 F-----GMOGEGLGYSWMSISPRVDRVKNKNVPRR-----PFGGSGVPR 1052
Db 1359 FVRVYASSDKTVEVKFGMKRSFVDSGEMS-----ESFPYTKALFAFEEDGVV 1408
```

```
QY 1053 EFTGQIIPST---PHELPGMG-----LSGSSSAVQE----- 1080
Db 1409 CFQGMHVQDTALIAPHQIOGCVVYISYLDSTHFPRPCLRTAVYHEILGYLEVYKLVVY 1468
QY 1081 -----HODD---TOHNOQDEMKNASHLO-----KTFDLNLSSECLTROSS 1119
Db 1469 TAIWACPPSEGDDYFHCCHPPQOKIPKPKRLOEWYKMKLDKAFABRIINDYKIDIFQAN 1528
QY 1120 TKQNTIDGCLPRDRTADVDVPLSNSSQNILVESNSSNKEQTAVEYKETAATILIREMK 1179
Db 1529 -----EDRLTSAKELPYFEGDFWPNVLEESIKELEQEERKEEASTAASETPE 1577
QY 1180 GTIADCKKPTSDWSLURKQVEGNEGROERNKNMNSIDYAIRRASISEISEAIKER--- 1236
Db 1578 GSQDSK-----NAKKKNKKTKNKK-----SSISRANKKKPSM 1611
QY 1237 -GNMNLAVRIKDFLER-----IVKDHGIDLEWLRESPPDKAKDYLLSIRGLGLKSVE 1289
Db 1612 PNYNDLSQKLYATMEKHKEVFFVIHLHAG---PVISTQPPIVDPDPLLS----- 1658
QY 1290 CVRLTLNLHNAFPVDNTNVR-----IAVRMGWVPLQPLPESLQLHLELYPVLESIQKFL 1344
Db 1659 -----CDLMDGRDAFLTLLARDKH-----EFSSLRRSK 1686
QY 1345 WPRICKLDQRTLYELHYQLITFGK---VFCTKSRPNACPMRCEGRHFASAYASAEALAL 1401
Db 1687 WSTLK-----MLVELHTQ---GQDREY-----TCNECKHHVETRMHCTVCEDYDLCI 1731
QY 1402 PAPEERLSATIPVPPSPFPAIPMIELPLPL-EKSLASGAPSNRENCEPIIEEPASP 1460
Db 1732 NCYNTKSHTH-----KMWKGLGLDDGSSQGEPOSK-----SP 1765
QY 1461 GQECTETI-----ESDIEDAYNEDPDEIPTIKLINIEQFGMTLREHMERMELOEG----- 1511
Db 1766 -QESRRLSIQRCISQVLHVACQCRNANGSLPSC-----QKMKRVVQHTKGCKRK 1812
QY 1512 -----DMSKALVAL-----HPTTTSIPTP---KLK-NISRLRTEHOVVELPDSHRLLDG 1556
Db 1813 TNGCPCVKQOLIAUCCYHAKHCQENKCPVFCNLKINVRQOQIQHCLQOQLMRMRMAT 1872
QY 1557 MDRK---EPDDPSPYLLAIWTPGE---TANSAPQPEQ 1587
Db 1873 MNTNRVQOQSLPSPTSAPPGTPTQOPSTPTQPPPAQ 1909
```

RESULT 10
US-08-194-468-2
; Sequence 2, Application US/08194468
; Patent No. 5750336
; GENERAL INFORMATION:
; APPLICANT: Montminy, Marc R.
; TITLE OF INVENTION: ASSAYS FOR THE IDENTIFICATION OF
; TITLE OF INVENTION: COMPOUNDS WHICH INHIBIT ACTIVATION OF CAMP AND MITOGEN
; TITLE OF INVENTION: RESPONSIVE GENES
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pretty, Schroeder, Brueggemann & Clark
; STREET: 444 South Flower Street, Suite 2000
; CITY: Los Angeles
; STATE: California
; COUNTRY: USA
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/194,468
; FILING DATE: 10-FEB-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Reiter, Stephen E.

```

; ; REGISTRATION NUMBER: 31.192
; ; REFERENCE/DOCKET NUMBER: P41 9672
; ; TELECOMMUNICATION INFORMATION:
; ; TELEPHONE: (619)-546-4737
; ; TELEFAX: (619)-546-9392
; ; INFORMATION FOR SEQ ID NO: 2:
; ; SEQUENCE CHARACTERISTICS:
; ; LENGTH: 2441 amino acids
; ; TYPE: amino acid
; ; TOPOLOGY: linear
; ; MOLECULE TYPE: protein
; ; US-08-194-468-2

```

Query Match 2.0%; Score 183; DB 1; Length 2441;
Best Local Similarity 17.5%; Pred. NO. 1.4e-05;
Matches 342; Conservative 222; Mismatches 685; Indels 70

84;

Qy	5	MDSAVNATEATEQNDGSRDVLFEOLNKT-----PQOKPSRKRKFMFKVVEGKPKRK	59
Db	287	MGATGVNQLASKOSVMVNSLPAPFTDIKNTVTVTPNMSQLQTSVGVPTQALATGPTAD	346
Qy	60	P--RKPAELPKVVVEGKPKRKPKAATQEKVSKETGSAKKKNLESAT-----KCPANVG	113
Db	347	PERKKLIQOOLVLLHLHAHKCORREOANGVEVRACSLPHCRTMKNVLNHHMTHCOAPKACQVA	406
Qy	114	DMSNKSPEVT--LKSCRK---ALNFDLENPCDARQGD-----SESEIVONSNGANSSEIRD	165
Db	407	HCASSRQIISHWNKCTRHDPCVCLPLKNASDRNQOTILGSPASGIONTIGS-----V	459
Qy	166	AIGGTNGSFLDSVSQIDTKN-----GLGAMNOPLEV-----	196
Db	460	GAGCQONATISNPNPDPDSMORAYAAALGLPMYNQOTOLQPVPCQQAQPPAHQOMRT	519
Qy	197	--SMGNOPDKLSTGAKLARDQOQDILLTRNQOQCFVATONTQFMEHQAWLQMNQLIG	254
Db	520	LNALGNPNMSPVAGG--ITTDQPPNLSISALPTSUGATN---PLMND-----G	564
Qy	255	FPGNOOPRMTIRNQOCPCLAMGNQOPMY-----LIGTPRPAL-----	291
Db	565	SNSGNIGSLSTIPTAAAPPSSGTVRKQWHEHVTDQLRSHLVHLVQAIPTTPDPAALKDRR	624
Qy	292	-----VSGNQQLG	299
Db	625	MENLVAYAKKVEGDMYESANSRDEYVHLLAEKIYKIQELEEKRRTRLHKOGILGNQPAL	684
Qy	300	GPQGNKRPIFLNHQTCLPAGNOLYGSPTDMHOLVMSTGGQHQGLLIKNOQPSGLTRGOOP	359
Db	685	PASSAQPPVTPPAQSVRPPNGPL---PLPVNRMYQSOG-----MNSFNPMSLGNVQLP	734
Qy	360	CVPLIDQOPATPKGFTHLNQMV--ATSMSS--PG--LRPHSQSQVPTTYLHVESVSRILNGT	415
Db	735	-----CAPMGPRAAAFPMHNSVQMNMSWASVPGMATSPSRMPPPN-----MMGT	777
Qy	416	TGTCQRSRAPYDSLQODIHQNKYILSHEISNGNCKKALPONSLSLTPPIMAKLEEARG	475
Db	778	HANNIMQAQPTQNF-----LPQN-----	796
Qy	476	SKROYHRAMGQTEKHDNLNLAQTAQSDQYVERHNSSTCVE-----YILDAAKTKIKQVQOE	530
Db	797	---QFPSSSGAMSVNSVGMQCPAAQAQGVSGQEPGAALPNLMLAPAQASQLCPVPVQGS	853
Qy	531	NLHGMPEVEIETDDP-----TDGARKKGNKTASISKGASK	565
Db	854	PLHPTPPASTAAGMPSLQHTAPGWTPPQAPATQPSPTPVSSGGTPTPTPGVSVAQAQT	913
Qy	566	GNSSPYKKAETAEKCKIVPKTPAKKGRAGRKKSVPPPAHASEIOLQMT-----PKPTLPS	620
Db	914	QSTPTVQAAQAQVTPQOPTVQ-----PPSVATPPQSSQQQPTPVHTQPTCTPLS	963
Qy	621	RSK-----PKGGRKSTQDSCKARGPSGELLQODSTAEIILYRMONLYLGDKREQEON	673
Db	964	QAAASIDNRVPTSTVTSAETSSQOQPDVPM-----EMKTEVOT-----DQAEPE-----	1011

Qy	674	AMVLYKGDGALVPYESKKRPKPYDIDDETTRIWNLLMGKDEKGEDEKDKKKKKWWE	733
Db	1011	-----PTES-KGEPRSEMMEED-----LQSSQVKEETDTTTEQKSEPMEV	1049
Qy	734	EERRVFRGRADSFIA RMHLVQGDRRFSPWKGSVDSVIGVFLTQNVNSDHLSSSAFMSLAA	793
Db	1050	ERKK-----PEYKVAKEEBEENSNDTASQSTSP	1078
Qy	794	RFPFKLSSSREDER-----NVRSVVDEPEGCIL-----NLNEIPSQOEKQVQHPDSMEV	842
Db	1079	SQPRKKIFPEELRQALMTLEALYRQDPESLPFRQPVDPQLGIPDYFDIVKVPMDLST	1138
Qy	843	--SGVDSGSKQE-----LRDCNSGIERFNFLKSIONLEEEVLSQDS	884
Db	1139	IKRKLDTOYQEPWQYVDDVRLMFNNAWLYNRKTSRV--YKFCSKLAEPFQEI-----	1190
Qy	885	FDAIFQSCRGVSGSCSKSDAEFPTTRCTKTVSGTSQSQVOTGSPNLSDEIC-----L	938
Db	1191	-DP-VMQS---LGYYCGRGYEFSPQTLCCYQKQLCTIPRDAAYSYQNRHYFCGKCFTEI	1245
Qy	939	QGNRPHLYEGSGDVOKQETTNVAOKKPDKLEKTMN-----	973
Db	1246	QG-ENTVL-----GDDPSQPTTIS--KQEFKKKKNDTLDPEFVDCKEGGRMKHQICVLH	1298
Qy	974	---WKDS-VC-----FQQRNDTNW-----QTTTPSSSY-----EQCATRQPHVLDIED	1012
Db	1299	YDIIWPSGFCVDCNLAKTGRPRKENKFSAKRLQTTTLGNHLEDRVKNKFLRQNHPEAGEV	1358
Qy	1013	F-----GMQGBGLGYSWMSISPRVDRVKNKNVPRR-----FFRGGSVPR	1052
Db	1359	FVRVASSDKTVEVKPGMKSRFVDSGEMS-----ESFPYRTKALFAFEEDIGVDV	1408
Qy	1053	EFTGQIIPST---PHELPMG-----LGSSSAVOE-----	1080
Db	1409	CFFGMHVODTALIAPHOIGCVVISYLDISIHFPRPRCLRTAVYHEILIGLEYVKKLWYV	1468
Qy	1081	---HQDD---TQHNQDENMKASHLQ-----KTFDLNLSSEECITROSS	1119
Db	1469	TAHIWACPSEGDDYTFCHCPDOKIPKPKRLQEWYKMLDKAFARIINDYKDIFKQAN	1528
Qy	1120	TKONITDGCPLPRDTAEDVVDPLSNNSNQNLIVESNSNKBQTAVEYKETNATILREMK	1179
Db	1529	-----EDRLTSAKELPYFEGDFWPNVLESSIKELQEEBEERKKEESTAASETPE	1577
Qy	1180	GTLADGKPTTSQWDSLRKDVNEGREGQERKNKMDSIDYEAIRRASISEISEATKER---	1236
Db	1578	SGQGDGK-----NAKKNNKNTKNK-----SSISRANKKKPSM	1611
Qy	1237	-GMNNMLAVRIKDFLER-----IVKDGGIDLEWLRESPPDKAKDYLLSIRGLGLKSVE	1289
Db	1612	PNVSNLDSQKLYATMEKHKEVFVTHLHAG---PVIQTQPIVDPDPLLS-----	1658
Qy	1290	CVRLTLHLNLAFFVDTNVGR-----IAVRMGVWVLOQLPESLQLHLLELYPVLESIOKFL	1344
Db	1659	-----CDLMDGRDAFLTLARDKHW-----EFSSLRRSK	1686
Qy	1345	WPRLCKLDQRTLHYELHYQLITGK---VPCITSRRPNCNACPMRGECRFPASAYASARAL	1401
Db	1687	WSTLC-----MLVELHTQ---GQDRFVY-----TCNECKHHVETRWHCITVCEDYDLCI	1731
Qy	1402	PAPERSLTSATIPVPSPFPVPAIPMIELPLPL-EKSLAGSAPNRENCPEIIEEPASP	1460
Db	1732	NCYNTKSHTH-----KWKWGLGLDDESSGQGFQSK-----SP	1765
Qy	1461	GOECTEIT-----ESDIEDAYNEDPDEIPTIKLNIQFGMTLREHMNMNMELOEQ----	1511
Db	1766	-QESRRLSIQRCIQSLSHVACQECNANCSLPSC-----QKMKRVVQHTKGCKRK	1812
Qy	1512	----DMSKALVAL-----HPTTTSIPPP---KUK-NISRLRTEHQVYELPDSHRLLDG	1556
Db	1813	TNGGCPYCKOLIAICCYHAKHQENKCPVPFCLNTIKHNVROOIQHICQOOLAMRRMRMAT	1872


```
Db 1053 EVG-----SSINEGSSDENIAELG-RNRGPKLNAMLRL---GVLPQEVYKQSLPGSNC 1103
Qy 692 KRPRKVIDIDETRIWNL-----LMKGDEKGD 721
Db 1104 KHPEIKQVEVEVQTVNTDPSYLI SDNLEQPMGSSHASQVSETPDDLDDGEIKEDT 1163
Qy 722 E--BKDKKKKWEERVF-----RG---RADSFIA RMHLVQGRFRSPWKGSVVDVSI 771
Db 1164 SFAENDIK-----ESSAVFSKVQSGELSRSPFTHTHLAQYRGA-----1206
Qy 772 GVFLTNVSHLSSAFMSLAARPPKLSRSDREVRNRSVVVEDPE-GCILNL-----N 825
Db 1207 -----KKLESS---EENLSS---EDELPCFCHLLFGKVN 1235
Qy 826 EIPSQWQKVQHPSPMEYSGVDSGSK-----OLRDCNSGIERFNFLKSIQNLVEEV 878
Db 1236 NIPS--OSTRH-STVATECSKNTENLLSLKNSLDCSNQVI-----LAKASQ---EHH 1284
Qy 879 LSSQDSFDPALFQSGRVGSCSKSDAEFFPTRCTKTYSQSVQGTGSPN-----LSD 934
Db 1285 LSEETKCSALFSS-----QCSELEDLTANTNTQDPFLIGSSKQMRHOSQGVGLSD 1337
Qy 935 EICQGNRPHLYGSGDVOKQETTNVAQKPDLEKTMNWKDSCVFCQPRNDTNWQTTPS 994
Db 1338 KELVSDDEE---RGTG---LEENN--QEEQSMDSNLGEAASGC-----ESET 1376
Qy 995 SSYEQCATRQPHVLIDIEDFGQEGELGYSNMSISPRVDRVKNKNVPRFRQGGSVVPREF 1054
Db 1377 SVSEDCS-----1383
Qy 1055 TGQIIPSTPHLPQMGGLSGSSSAVQEHQDDT-QHNOQDEMKNKASHLOKTFDLNLSSEEC 1113
Db 1384 -----GLSSQSDILTTQORDTMQHN-----LIKQOEMAELEAVLEQH 1421
Qy 1114 LTROSSTKONTDCLPRDRTAEDVDPLSNSSLIQNLVSNSSKEQTAVEKETNAT 1173
Db 1422 GSQPSNYSPT-----ISDSSALEDLRNPEQSTSEKAVLTSOKSEYP 1464
Qy 1174 ILRMKGTGLADGKPTQOWDSLKRDVEGNEGRQERNKNMDSIDYEAIRASISEAISEAI 1233
Db 1465 ISQNEGLSAD-KREVSADSTSKNKE--FCOVERSSPKCPSLD-----DRWMHSCSGSL 1517
Qy 1234 KERGM-NMMLAVRIKDFLERIVKDHGGIDL 1262
Db 1518 QNRNYPQSEELIKVVDVEEQLEESGPHDL 1547
```

RESULT 13

```
US-08-798-691-6
; Sequence 6, Application US/08798691
; Patent No. 5750400
; GENERAL INFORMATION:
; APPLICANT: Murphy, Patricia D.
; APPLICANT: Allen, Antonette C.
; APPLICANT: Alvares, Christopher P.
; APPLICANT: Critz, Brenda S.
; APPLICANT: Olson, Sheri J.
; APPLICANT: Schelter, Denise B.
; APPLICANT: Zeng, Bin
; TITLE OF INVENTION: Coding Sequences of the Human
; TITLE OF INVENTION: BRCA1 Gene
; NUMBER OF SEQUENCES: 72
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ONCORMED
; STREET: 200 Perry Parkway
; CITY: Gaithersburg
; STATE: MD
; COUNTRY: USA
; ZIP: 20877
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER:- IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
```

```
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/798,691
; FILING DATE: 12-Feb-97
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Thomas Gallegos
; REGISTRATION NUMBER: 32,692
; REFERENCE/DOCKET NUMBER: PA-0054CIP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 301-527-2051
; TELEFAX: 301-208-6997
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1863 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; STRAIN: BRCA1
; POSITION IN GENOME:
; CHROMOSOME/SEGMENT: 17
; MAP POSITION: 17q21
; US-08-798-691-6
```

Query Match 2.0%; Score 182; DB 1; Length 1863;

Best Local Similarity 18.4%; Pred. No. le-05;

Matches 293; Conservative 195; Mismatches 494; Indels 608; Gaps 72;

```
Qy 20 DGSRODVLER---DLNKTPOQKPSKRKRKFMKPVVVEGPKR-----58
Db 219 DSAKAKACEFSETDVTTEHHQPSNNDLNTTEKRAARHPEKYGQSSVSNLHVEPCGNT 278
Qy 59 -----KRPKPAELPKVVVEGKP-----KRPKPAATQEKVKSETGSAKK 98
Db 279 HASSLQHENSLLLTDRMNMVKAFCNKSQKPGCLARSQHNRWAGSKETCNDRTTPSTEK 338
Qy 99 K-----NLKESATKKPANVGDMNSKSEVTLKSCRKALN-----FDLEN 137
Db 339 KVDLNADPLCERKEWNKQKLPCESENPRDTEDEV---EWITLNSSIQKVNEMFSRDELLG 394
Qy 138 PGDARQDSESEI-----VQNSSGANS-----FS 161
Db 395 SDDSHDGESNAKADVLDVLNEVDEYSGSSEKIDLLASDPHEALICKSERVHSKSVES 454
Qy 162 EIRDAIGGTN-----GSFLDSVSQIDK---TNGLGAMNQPLEVSMG 199
Db 455 NIEDKIFGKTYRKASUPNLSHVTENLIIGAFVTEPQIIQERPLTNKLRKRRLP---TSG 511
Qy 200 NOPDKLSTGAKLARDQOPDLLTR--NQCCQFPVATQNTQPMENQOAWLQKNQIGLGFPP 257
Db 512 LHPEDFIKKADLAVOKTPEMINOGTQNTQEQNGVQVMNTTNSGHENKTKGDSIQNE-----565
Qy 258 GNOQPRMTIRNOOPCLAMGNOQPMYLLIGTPRALVSGNOOLGGPGQGNKRPIFLA-HOTCL 316
Db 566 -----KNPNPIESLEKESAFKTAEPISISSIS-NMELE-----LNIHNSKA 605
Qy 317 PAGNQLYGSPTDMH-----QLVMS-----TGGOQHGLLIKNQOPGSLIRG 356
Db 606 PKNRLRRKSSTRHHALELVVSRNLSPNPCTELQIDSCSSSEIEIKKKYNQMP---VRH 662
Qy 357 QQPCVPLIDQOPAT-----PKGET-----HLNQMV 381
Db 663 SRNLQLMEGKEPATGAKKSNKPNQTSKRHSDTTFPELKLTNAPSGFSCNTSELKEFV 722
Qy 382 ATSM-----SSPGLRPHSQSQVP-TTYLHVES 407
Db 723 NPSPREEKEKLETVKVSNNAEDPKDMLMSGERVLOTERSVSSSSISLVPGDTGYGTQES 782
Qy 408 VSRILNCTTGTQCR-----SRAPAYDSLQODTH-----QGNKYILSHSISNGNG 451
```



```
Db 783 ISLEVTGLKAKTEPNKVCYSCAAAFENPGLIHGCKDNRNDETEGPKYPLGHEV----- 837
Qy 452 CKKALPONSSLPPTIMAKLEA-----RGSKRQ-----YHRA 483
Db 838 -----NHRETSI--EMESELDAQYLONTFKVSRQSFALFSPNGNAEECATSAH 888
Qy 484 MGOTEKHLNLAQIAQODVERHNSSTCVEYLDAAKTKIKYVQENL-HGMPPEVIEI 542
Db 889 SGLKQSPKVTPECEKEENQGNES-----NIKPVQTVNITAGFP--VVGQ 934
Qy 543 EDDPTDGR---KCKNTASIS---KGASKNSSPVKTAKEKCIVP-----KTPA 587
Db 935 KDKFVDNAKSIKGGSRFCLSSQFRGNETGLITPNKKGHLQNPYRIPPLPKFSFVKTC 994
Qy 588 KKGAGRKKSVPHPAHASEIQLOMTPPKT--PLSRSKP-----KKGKRSIODSGKARG 640
Db 995 KKNLL--EENFEHSMSPEREMGNENTPSTVTSIRNIRENFVKGASSNINEVGSTN 1052
Qy 641 PSGELLQDSIAEIIYRMQNL--LGDKEREQONAMVLYKGDAIYVYESKK----- 691
Db 1053 EVG-----SSNETGSSDENITQAEIG--RNRGPKLNAMRL--GVLQPEVYKQLPGSNC 1103
Qy 692 RKRPKYDIDDETTRINL-----LMKGDEKEGD 721
Db 1104 KHPEIKKQEEYEVQTVNTDFSPYLISDNLEQPMGSSHASQVCSETPDDLLDGEIKEDT 1163
Qy 722 E--EKDKKKKEKWEERRVF-----RG---RADSFIAHMLVQCDRFRSPWKGSVVDVI 771
Db 1164 SFAENDIK-----ESSAVSKSVQRELSRSPFTHLAQYRRGA----- 1206
Qy 772 GVFLTQVNSHLSSAFMSLAARPPPKLSSREDERNVRSVVVEDPE--GCILNL-----N 825
Db 1207 -----KKLESS---EENLSS---EDELPCFQHLLEKVN 1235
Qy 826 EIPSWQEVQHPSDMEVSGVDSGKE-----QLRCSNSGIERFNFLEKSIQNLSEEV 878
Db 1236 NIPS--QSTRH--STVATECLSKNTEENLLSLKNSLDCSNQVI-----LAKASQ---EHH 1284
Qy 879 LSSQSDPDPAIFQCGRVGSCSKSDAEPPTTCETKTVSGTSQSQTGSPN-----LSD 934
Db 1285 LSEETKCSASLFSS-----QCSELEDLTANTQDPFLIGSSKQMRHOSQGVGLSD 1337
Qy 935 EICLOGNERPHLYGSGVQKQETTNVAQKPDLEKTMNMKDSVCFQOPRNDTNMQTTPS 994
Db 1338 KELVSDDEE---RGTG---LEENN--QEEQSDMSNLGEAASGC-----ESET 1376
Qy 995 SSVBQCATRQPHVLVDIEDFGMGGLGYSNMSISPRVDRVKNKNVPRFRFGQSVPRF 1054
Db 1377 SVSEDCS----- 1383
Qy 1055 TGOIIPSTPHELPGMGLSGSSSAVQEHQDDT-QHNOODEMKNKASHLOKTFDLNLSSEEC 1113
Db 1384 -----GLSSQSDILTTQORDTMQHN-----LIKLOEAMAELEAVLEQH 1421
Qy 1114 LTROSSTKQNTDGLCLPRDRTAEDVDPLNSNLSQNLVSNKQETQAVEYKETNAT 1173
Db 1422 GSQSNVYPSI-----ISDSSALEDLRNPQESTSEKAVILTSQKSSEYP 1464
Qy 1174 ILREMKGTLDGKKKPTSQWDSLRKDXVEGNEGRQERNKNWDSIDYEAIRASISEISEAI 1233
Db 1465 ISQNPGLSAD-KFEVSADSSSTKNKE--PCVERSFPKCPSLD---DRWMHSCSGSL 1517
Qy 1234 KERGM--NMMLAVRIKDFLERIVKDHGGIDL 1262
Db 1518 QNRNYPQEEELIKVVDVEEQOLESQPHDL 1547
```

RESULT 14

US-08-825-487A-2

; Sequence 2, Application US/08825487A

; Patent No. 6048689

; GENERAL INFORMATION:

```
APPLICANT: Murphy, Patricia D.
TITLE OF INVENTION: White, Maiga B.
METHODS FOR IDENTIFYING VARIATIONS IN POLYNUCLEOTIDE SEQUE
NUMBER OF SEQUENCES: 110
CORRESPONDENCE ADDRESS:
ADDRESSEE: Howrey & Simon
STREET: 1299 Pennsylvania Avenue., N.W.
CITY: Washington,
STATE: DC
COUNTRY: USA
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/825,487A
FILING DATE: 28-MAR-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US98/060002
FILING DATE: 26-Mar-1998
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Albert P. Halluin
REGISTRATION NUMBER: 25,227
REFERENCE/DOCKET NUMBER: 05371.0012.999
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-463-8100
TELEFAX: 650-463-8400
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1863 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: not relevant
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
STRAIN: BRCA1
POSITION IN GENOME:
CHROMOSOME/SEGMENT: 17
MAP POSITION: 17q21
US-08-825-487A-2
```

Query Match 2.0%; Score 182; DB 3; Length 1863;

Best Local Similarity 18.4%; Pred. No. 1e-05;

Matches 293; Conservative 195; Mismatches 494; Indels 608; Gaps 72;

```
Qy 20 DGSRODVLEF---DLNKTPOQKPSKRRKRMKMPKVVVEGPKPKR----- 58
Db 219 DSAKKAACEFSETDVTTEHQPSNNDLNTEKRAAEHRHPEKYOGSSVSNLHVPCGTNT 278
Qy 59 -----KRPKPAELPKVVVEGKP-----KRPKPAATQEKVKSKETGSAKK 98
Db 279 HASSLOHENSLLLTDKRMNVKAEFCNKSQPCQLARSOHNRWAGSKETCNDRTTPSTEK 338
Qy 99 K-----NIKESATKKPANVGDMKSNKSEVTLKSCRKALN-----FDLEN 137
Db 339 KVDLNADPLCERKEWNKQKLPCESENPRDTEV-----PWITLNSSIQKVNWFNFRSDELLG 394
Qy 138 PGDARQDSESEI-----VQNSSGANS-----FS 161
Db 395 SDDSHDGESESNKAVADVLDVLNEVDYESSGSEKIDLLASDPHEALICKSERVSKSVES 454
Qy 162 EIRDAIGTGN-----GSFLDSVSGIDK---TNGLGAMNQPLEVSMG 199
Db 455 NIEDKIEGKYRKASLPNLSHVNTENLIIGAFVTEPQIIQERPLTNKLRKRRP---TSG 511
Qy 200 NQPKLSTGAKLARDQOPDLLTR--NOOCOPPVATQNTQPFMENOQAWLQMKNLGIFPF 257
```

Db 512 LHPEDFIKKADLAVQKTPPEMINOQTQTEQNGQVMNITNSCHENKTKGDSIQNE----- 565
QY 258 GNOQPRWTIRNOOCLAMGNOQPMYLGTRPALVSGNQOLGGPOGNKRPIFLN-HOTCL 316
Db 566 -----KNPNPIESLEKESAFKTAEPISISSIS-NMELE-----LNIHNSKA 605
QY 317 PAGNQLYGSPTDMH-----OLVMS-----TGGOQHGLLIRKNOQPGSLIRG 356
Db 606 PKKNRLARKSSTRIHHALELVVSRNLPNPPNCTELQIDSCSSSEIEIKKKYNQMP---VRH 662
QY 357 QOPCVPLIDQOPAT-----PKGFT-----HLNQMV 381
Db 663 SRNLQLMGKREPATGAKSKNPNQTSKRHSDTTFPELKLTNAPGSPFTKGSNTSELKEFV 722
QY 382 ATSM-----SSPGLRPHSOSQVP-TTYLHVES 407
Db 723 NPSLPREKEKEKLETVKVSNNADPKDMLSGERVLQTERSVESSISLVPGTDYGTQES 782
QY 408 VSRILNCTTTCQCR-----SRAPAYDSLOQDIH-----QGNKYILSHEISNGNG 451
Db 783 ISLLEVSTLCKAKTEPNKCVSOCAAFENPKGLIHGCSKDNRNDETEGPKYPLGHEV----- 837
QY 452 CKKALPONSLPTPIIMAKLEEA-----RGSKRQ-----YHRA 483
Db 838 -----NHSRETSI--EMBESELDQYLOLQNTFKVSRKRQFALFNSPGNAEECATPSAH 888
QY 484 MGOTEKHDLNLAQIAQSDQVERHNSSTCVYELDAAKTKIKQVVOENL-HGMPPEVIEI 542
Db 889 SGSLKQSPKVTPECEKEENQGNES-----NIKPVQTVNITAGFP--VVGQ 934
QY 543 EDDPTDGAR---KGKNTPASIS---KGASKGNSSPVKTAKEKCIVP-----KTPA 587
Db 935 KDRPVDNAKCSIKGGRSFCFLSSQFRGNETGLITPNKHGLLONPVPPLFPKSFVKTKC 994
QY 588 KKGKGRKKSVPVPAHASETLQWPTPKT--PLSRSKP-----KGKGRKSQDSCKARG 640
Db 995 KKNLL--EENFEHMSPEPEMGNENIPSTVSTISRNIRENFVKGASSNINEVGSSTN 1052
QY 641 PSGELLCQDSTAETIYRMQNLV--LGSKEREQONAMVLYKGDGALVPYESKK----- 691
Db 1053 EVG-----SSINEIGSSDENIQALG-RNRGPKLNLMRL--GVLOPEYVKQSLPQSN 1103
QY 692 KRPKPKVIDDETTRIWL-----LMGKGRDEKEGD 721
Db 1104 KHPEIKKQYEEVVQTVNTDFSPYLSIDNLEQPMGSSHASQVCSETPDDLDDGETKEDT 1163
QY 722 E--EKDKKKKEKWEERRVE-----RG--RADSFARMHLVGDREFPWKGVSVVDVI 771
Db 1164 SFAENDIK-----ESSAVFSKSVQRGELSRSPFTHTHLAOGYRGA----- 1206
QY 772 GVFLTONVSDHLSSAPMSLAARFPKPLSSSRREDERNVRSVVVEDPE-GCILNL-----N 825
Db 1207 -----KLESS-----EENLSS--EDELPCFOHLLFGKVN 1235
QY 826 EIPSWQKVOHPSDMEVSGVDSGKE-----OLRDCNSNGIERFNFLEKSIQNLSEEV 878
Db 1236 NIPS--QSTRH-STVAPECLSKNTEENLLSLKNSLNCNQVI-----LAKASQ---EHH 1284
QY 879 LSSQDSFDPALFQSGCRGVGSCSKSDAEPPTTCETKTVSGTSSQVQGTGSPN---LSD 934
Db 1285 LSEETKCSASLFS-----QCSLEDLTANTQDPFLIGSSKQMRHOSQGVGLSD 1337
QY 935 EICLQGNRPHLYEGSDVQKQETTNVAQKPKDLEKTMNKNKDSVCFQGPQRNDTNWQTTPS 994
Db 1338 KELVSDDEE---RGTG---LEENN--QEEQSMDSNLGEAASGC-----ESET 1376
QY 995 SSVTEQCATRPHVLIDEDFCMQGEGLOYGYSWMSISPRVDRVKNKNVPRFRFGQSVVPREF 1054
Db 1377 SVSDEDCS----- 1383
QY 1055 TGOIIPSTHELPGMGLGSSSAVQEHQDDT-OHQOQDENMKASHLOKTFDLNLSSEEC 1113
Db 1384 -----GLSSQSDILITQOORTMOHN-----LIKQOEMAELEAVLEQH 1421

QY 1114 LTRQSSSTKQNTDGLCLPRDRTAEDVVDPLSNNSLQNLVSNSSNKEQTAVYKKTNAT 1173
Db 1422 GSOPSNYSYFI-----ISDSSALEDLRNPEQSTSEKAVLTQSKSSEYP 1464
QY 1174 ILREMGKTLADGKKPTSQWDSLRKDVNEGROERKNNKNNMSIDYAIRRASISEISEAI 1233
Db 1465 ISQNPGLSAD-KFEVSADSSSTSKNE--PGVERSSPSKCPSLD---DRWYMHSCGSL 1517
QY 1234 KERGM-NMMLAVRIKDFLERIVKDHGIDL 1262
Db 1518 QNRNYSQEBELIKVVDVEEQOLEESGPHDL 1547
RESULT 15
US-08-825-487A-6
; Sequence 6, Application US/08825487A
; Patent No. 6048689
; GENERAL INFORMATION:
; APPLICANT: Murphy, Patricia D.
; APPLICANT: White, Marga B.
; TITLE OF INVENTION: METHODS FOR IDENTIFYING VARIATIONS IN POLYNUCLEOTIDE SEQUE
; NUMBER OF SEQUENCES: 110
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Howrey & Simon
; STREET: 1299 Pennsylvania Avenue., N.W.
; CITY: Washington,
; STATE: DC
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/825,487A
; FILING DATE: 28-Mar-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US98/060002
; FILING DATE: 26-Mar-1998
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Albert P. Halluin
; REGISTRATION NUMBER: 25,227
; REFERENCE/OCCKET NUMBER: 05371.0012.999
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-463-8100
; TELEFAX: 650-463-8400
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1863 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; STRAIN: BRCA1
; POSITION IN GENOME:
; CHROMOSOME/SEGMENT: 17
; MAP POSITION: 17q21
US-08-825-487A-6

Query Match 2.0%; Score 182; DB 3; Length 1863;
Best Local Similarity 18.4%; Pred. NO. le-05;
Matches 293; Conservative 195; Mismatches 494; Indels 608; Gaps 72;

QY 20 DGSRODVLEP---DLNKTPOQKPSKRKRKFMKVYVVGKPKR----- 58
Db 219 DSARKACESEYDVTNTEHHQFSNNDLNTERAAARHPKQYQGSVSNLHVPCGTNT 278

Qy 59 -----KPRKPAELPKVVVEGRP-----RRKPKAAATQEKVSKETGSACK 98
Db 279 HASSLOQHENSLLTLTKDRMVEKAEFCNKSQKQGLARSHQHNRAWSKETCNDRRTPSTEK 338
Qy 99 K-----NLKESATKPPANVGMNSKSPVTLKSKRKALN-----FDLEN 137
Db 339 KVDLNAADPLCERKEWKKQLPCSEPRTEDEY-----PWITLSSIQKYNWFMSRDELLG 394
Qy 138 PGDARQGSSESI-----VONSSGANS-----FS 161
Db 395 SDDSHDGESENAKVADVDLVNVEDEYSGSEKIDLLASDPHEALICKSERVHSKSVES 454
Qy 162 ETRDAIGTN-----GSFLDSVSQIDK-----TNGLGAMNQPLEVSMG 199
Db 455 NIEDKIFGKTYRKASLPNLHVTEENLIIGAFVTEPOIIQERPLTNKLRKRPP-----TSG 511
Qy 200 NOPDKLSTGAKLARDQOQDPLLTR--NOQCQFPVATQNTQFPMENOQAWLQMKNLQIGPFF 257
Db 512 LHPEDFIKKADJAVOKTPEMINQGTQEQNGQVNMNITNSGHENKTKGDSIQNE-----565
Qy 258 GNOQPRMTRINQOPLAMGNOQPMYLIQTPRPALVSGNOOLGGPOGKNRPFLN--HOTCL 316
Db 566 -----KNPNPTESLEKESAFKTAEPITSSIS--NWELE-----LNIHNSKA 605
Qy 317 PAGNOLYGSPTDMH-----OLVMS-----TGGQOHLIKKNOQPGSLIRG 356
Db 606 PKKNRLRRKSSRTHIALELVVSRNLSPNCTELQIDSCSSSEIEKKKKYNQMP--VRH 662
Qy 357 QQPCVPLIDQOPAT-----PKGFT-----HLNOMV 381
Db 663 SRNLQMEGKEPATGAKKNKPNQEQTSKRHSDTPELKLNTNAPGSFTKCSNTSIELKEFV 722
Qy 382 ATSM-----SSPGLRPHSQSOVP--TTVLHVES 407
Db 723 NPSPREKEKEKLETVKYSNNAEDPKDMLSGERVLOQTERSVESSISLVPGTDYGTQES 782
Qy 408 VSRILNGTGTQCR-----GRAPAYDSLOQDIH-----QGNKYILSHSISNG 451
Db 783 ISLLEVSTLGRAKTEPNKCVSOCAAFENPKGLIHGCSKDNRRNDEGPKVPLGHEV-----837
Qy 452 CKKALPONSSUPTPIMAKLEA-----RGSKRQ-----YHRA 483
Db 838 -----NHSRETST--EMEESELDAQYLQNTFKYKRSQFALFSPNGNAEBECATFSAH 888
Qy 484 MGOTEKHDLNLAQQAQODYVERHNSSCTVEYLDAAKTKIKQVVOENL--HGMPEVIEI 542
Db 889 SGLKKQSPKVTFECEQKEENQGNES-----NIKPVQTVNITAGFP--VYQ 934
Qy 543 EDDPTDGAR--KGKNTASIS--KGASKGNSPVYKKTAEKEKCIVP-----KTPA 587
Db 935 KDKPVDNAKCSIKGSRFCLSQPRNETGLITPNKHGLLQNPYRIPPLFPKISFVKTKC 994
Qy 588 KKGRAGRKSVPPPAHASEIQWLQWTPPKT--PLSRXP-----KGGRKSQIDSGKARG 640
Db 995 KKNLL--EENFEHSMSPERMGNEINPSTVSTISRNIRENVFKGASSNINEVGSSTN 1052
Qy 641 PSGELLCODSTAEIIRYMNLY--LGDKEREOQNAWLYKGDGALVPYESKK-----691
Db 1053 EVG-----SSINEIGSSDENIQAELG--RNRGPKLNAMELRL--GVLQPEVYKQSLPGSNC 1103
Qy 692 RKPRPKVDIDDETTRIWNL-----LMGKGDEKRGD 721
Db 1104 KHPEIKKQYEVEVQVTVNTDFSPYLI SDNLEQPMGSSHASQVCSETPDLDLDDGEIKEDT 1163
Qy 722 E--EKDKKKEKWEERRVF-----RG--RADSFIAHMLVQGRFRFPWKGVSVDVSI 771
Db 1164 SFAENDIK-----ESSAVFSKVQGELSRSPPFTTHLAQGYRGA-----1206
Qy 772 GVFLTQNVSDHLSSSAFMSLAARPPKLSRSSREDERNVRSVVVDEPPE--GCILNL-----N 825
Db 1207 -----KKLESS-----EDELPCFOHLLFGKVN 1235

Qy 826 EIPSQEKVQHPMDMEVSGVDGSKSE-----QLRDCSNSGIERFNFLEKSIQNLBEEV 878
Db 1236 NIPS--QSTRH--STVATECLSKNTENLLSLKNSLNDCSNQVI-----LAKASQ-----EHH 1284
Qy 879 LSSQDSFDPALFQSCGRVSGSCSKSDAEFFPTTRCTETKTVSGTSQSVQGTGSPN-----LSD 934
Db 1285 LSEETKCSASLFS-----QCSELEDLTANTNTQDPFLIGSSKQMRHQSOGVGGLSD 1337
Qy 935 EICLOGNERPHLYEGSGDVQKQETTNAQKPKDLCKTNWKKDSVCFGQPRNDTNMQTTPS 994
Db 1338 KELVSDDEE-----RGTG-----LEENN--QEQSDMSNLGEAASGC-----ESET 1376
Qy 995 SSYEOCATRQPHVLDIEDFGMQGEGLYGWSMISPRVDRVKNNKNNVPRFRFGGSGVPREF 1054
Db 1377 SVSEDCS-----GLSSQSDILTTQQRDTMQHN-----LIKLOQEMAELEAVLEOH 1421
Qy 1055 TQOIIPSTPHELPGMGLSGSSSAVQEHODDT--OHNQODEMKNKASHLOKTFDLDLLNSSEEC 1113
Db 1384 -----GLSSQSDILTTQQRDTMQHN-----LIKLOQEMAELEAVLEOH 1421
Qy 1114 LTROSSTKONITDGCPLPRDRTAEDVVDPLSNSSNQILVSNSSNKQBTAVYKTNAT 1173
Db 1422 GSQPSNSYPSI-----ISDSSALEDLRNPEQSTSEKAVLTSOKASSEYP 1464
Qy 1174 ILREMGTLADGKPKTSQWDSLRLKDVNEGRQERKNKNMDSIDYEAIRRASISISEIAI 1233
Db 1465 ISQNPGLSAD--KFEVSADSTFSKNKE--PGVERSSPKCPSLD-----DRWYMHSCSGSL 1517
Qy 1234 KERGM--NNMLAVRIKDFLERIVKDHGGIDL 1262
Db 1518 QNRNYPQOEELIKVYDVEEQLEESGPHDL 1547

Search completed: September 27, 2002, 12:55:48
Job time: 9577 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 27, 2002, 10:26:26 ; Search time 70.37 Seconds
(without alignments)
2360.927 Million cell updates/sec

Title: US-09-840-743-2
Perfect score: 9089
Sequence: 1 MOSIMSSAVNATEQND.....PRPLMARLHPASKLNKNT 1729

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR71.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	3949	43.4	1017	2 T48452	hypothetical prote
2	2863	31.5	555	2 T48453	hypothetical prote
3	1559.5	17.2	1207	2 D84781	hypothetical prote
4	1260	13.9	234	2 T48454	hypothetical prote
5	709	7.8	917	2 T03430	hypothetical prote
6	278	3.1	5327	2 T13564	microtubule-associ
7	246.5	2.7	2897	2 B48666	cell proliferation
8	246.5	2.7	3256	2 A48666	cell proliferation
9	240.5	2.6	2938	2 T30249	hypothetical prote
10	233.5	2.6	2218	2 B48683	endonuclease III -
11	232	2.6	259	2 D75275	probable heat choc
12	219.5	2.4	1871	2 D96796	hypothetical prote
13	217.5	2.4	1490	2 T20513	glucan 1,4-alpha-g
14	217	2.4	2022	2 T48818	hypothetical prote
15	211	2.3	1188	2 T05324	hypothetical prote
16	210.5	2.3	1805	2 A34736	nestin - rat
17	208.5	2.3	3924	2 S37431	ankyrin 2, neurona
18	207	2.3	2331	2 T25410	hypothetical prote
19	206	2.3	3187	2 JC5837	364K Golgi complex
20	205.5	2.3	2562	2 T14266	Xin protein - chic
21	205	2.3	1983	2 T00385	KIAA0624 protein -
22	204.5	2.2	2717	2 A34203	DNA-binding protei
23	197.5	2.2	1145	2 T18235	transcription acti
24	197	2.2	1366	2 B86292	F7H2.12 protein -
25	196.5	2.2	3225	2 I52300	giantin - human
26	196	2.2	3488	2 T34418	hypothetical prote
27	192.5	2.1	2954	2 T14156	kinesin-related pr
28	192.5	2.1	3259	1 A56539	giantin - human
29	192	2.1	1888	2 T14273	zinc finger protei

ALIGNMENTS

RESULT 1

T48452
hypothetical protein T32M21.160 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 20-Apr-2000
C:Accession: T48452
R:Bevan, M.; Terry, N.; Ardiles, W.; Buysshaert, C.; Dasseville, R.; De Clerck, R.;
ewes, H.W.; Rudd, S.; Lemcke, K.; Mayer, K.F.X.
submitted to the Protein Sequence Database, March 2000
A:Reference number: Z24487
A:Accession: T48452
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1017 <BEV>
A:Cross-references: EMBL:AL162875
A:Experimental source: cultivar Columbia; BAC clone T32M21
C:Genetics:
A:Map position: 5
A:Introns: 167/1; 874/1
A:Note: T32M21.160

Query Match	43.4%	Score	3949;	DB	2;	Length	1017;
Best Local Similarity	99.9%	Pred. No.	1.1e-206;				
Matches	755;	Conservative	1;	Mismatches	0;	Indels	0;
QY	1	MOSIMSSAVNATEQNDGSRQDVFLEFDLNTPOQKFSKRKRKMPKVVVEGPKRKP	60				
Db	235	MOSIMSSAVNATEQNDGSRQDVFLEFDLNTPOQKFSKRKRKMPKVVVEGPKRKP	294				
QY	61	RKPAELPKVVVEGPKRKRKMPKVVVEGPKRKRKMPKVVVEGPKRKRKMPKVVVEGPKRKRKMP	120				
Db	295	RKPAELPKVVVEGPKRKRKMPKVVVEGPKRKRKMPKVVVEGPKRKRKMPKVVVEGPKRKRKMP	354				
QY	121	EYTLSCRKALNFLENPDGARGDSEIVQNSGANSFSEIRDAIGGTNGSFLDSVSQ	180				
Db	355	EYTLSCRKALNFLENPDGARGDSEIVQNSGANSFSEIRDAIGGTNGSFLDSVSQ	414				
QY	181	IDKTNGLGAMNQLGIFPGNQPMIRNQOCPCLANGNQPMYLTGTPRALVSGNQQLGG	240				
Db	415	IDKTNGLGAMNQLGIFPGNQPMIRNQOCPCLANGNQPMYLTGTPRALVSGNQQLGG	474				
QY	241	NQOAWLQKMLNQLGIFPGNQPMIRNQOCPCLANGNQPMYLTGTPRALVSGNQQLGG	300				
Db	475	NQOAWLQKMLNQLGIFPGNQPMIRNQOCPCLANGNQPMYLTGTPRALVSGNQQLGG	534				
QY	301	PGNKRPIFLNHQTCPLPAGNOLYGSPTDMHQLVMSGTGGQHQGLLKNQOGLSLIRGQGPC	360				
Db	535	PGNKRPIFLNHQTCPLPAGNOLYGSPTDMHQLVMSGTGGQHQGLLKNQOGLSLIRGQGPC	594				
QY	361	VPLIDQOQATPKGFTHLNQVATSMSSPGLRPHSOSQVPTTYLHVESVSRILNGTTGTCQ	420				

Db 595 VPLIQOAPATPKGFTHLNOMVATSSSPGELRPHSOSVPTTYLHWESVRLINGTTGTQC 654
 QY 421 RSRAPAYDSLQODIHQGNKYLISHEISNGGCKKALPONSILPTPIMALEFARSGKROY 480
 Db 655 RSRAPAYDSLQODIHQGNKYLISHEISNGGCKKALPONSILPTPIMALEFARSGKROY 714
 QY 481 HRAMQTEHDLNLAQOIAQOSODVERHNSSTCEVYLDAAKTKIKOVQENJHGMPEVI 540
 Db 715 HRAMQTEHDLNLAQOIAQOSODVERHNSSTCEVYLDAAKTKIKOVQENJHGMPEVI 774
 QY 541 EIEDPTGARGKNTASISKSGKNSPVKTAKEKCTYPKPAKGRGKRSKSVPR 600
 Db 775 EIEDPTGARGKNTASISKSGKNSPVKTAKEKCTYPKPAKGRGKRSKSVPR 834
 QY 601 PAHASEIOLMOTPPKPTPLSRKSPKGRKSIODSGKARGPSGELLCDPSIAEIIYRMON 660
 Db 835 PAHASEIOLMOTPPKPTPLSRKSPKGRKSIODSGKARGPSGELLCDPSIAEIIYRMON 894
 QY 661 LYLGDKEKEQONANVLYKGGDALVPESSKRRKPRKYDIDETTRINMLMGKDEKEG 720
 Db 895 LYLGDKEKEQONANVLYKGGDALVPESSKRRKPRKYDIDETTRINMLMGKDEKEG 954
 QY 721 DEEKDKKKKKWEEERRVFRGRADSFIAHMLVQGD 756
 Db 955 DEEKDKKKKKWEEERRVFRGRADSFIAHMLVQGE 990

RESULT 2

148453
 hypothetical protein T32M21.170 - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 20-Apr-2000 #sequence_rev1sion 20-Apr-2000 #text_change 20-Apr-2000
 C:Accession: T48453
 R:Bevan, M.; Terry, N.; Ardiles, W.; Buysshaert, C.; Dasseville, R.; De Clerck, R.; De
 ewes, H.W.; Rudd, S.; Lemcke, K.; Mayer, K.F.X.
 submitted to the Protein Sequence Database, March 2000
 A:Reference number: Z24487
 A:Accession: T48453
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-555 <BEV>
 A:Cross-references: EMBL:AL162875
 A:Experimental source: cultivar Columbia; BAC clone T32M21
 C:Genetics:
 A:Map position: 5
 A:Introns: 469/3; 496/2; 524/3
 A:Note: T32M21.170

Query Match 31.5%; Score 2863; DB 2; Length 555;
 Best Local Similarity 100.0%; Pred. No. 5e-148;
 Matches 549; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 785 SSAPFSLAARPPKLTSSRREDERNRVSIVVDEPCCILNLNLEIPSMOKVQHPMSMEVSG 844
 Db 7 SSAPFSLAARPPKLTSSRREDERNRVSIVVDEPCCILNLNLEIPSMOKVQHPMSMEVSG 66
 QY 845 VDSGSKELRDCNSNGIERFNFLEKSIONLEEEVLSODSDFPAFOCGRGVSCSGCS 904
 Db 67 VDSGSKELRDCNSNGIERFNFLEKSIONLEEEVLSODSDFPAFOCGRGVSCSGCS 126
 QY 905 DAEPTTRCETKTVSGTSQVOTGSPNLSDICLOGNERPRLYEGSGDVQKQOETTNAOK 964
 Db 127 DAEPTTRCETKTVSGTSQVOTGSPNLSDICLOGNERPRLYEGSGDVQKQOETTNAOK 186
 QY 965 KPDLEKTMNMDVSCFQPRNDTMMQTTTPSSSYECACATROPHVLDIEDFGKQGBGLGYSW 1024
 Db 187 KPDLEKTMNMDVSCFQPRNDTMMQTTTPSSSYECACATROPHVLDIEDFGKQGBGLGYSW 246
 QY 1025 MSISPRVDVAVKNNVPRRFFQGGSVPRREFTGQIIPSTPHLPGMGSLGSSSAVQEHODD 1084
 Db 246 MSISPRVDVAVKNNVPRRFFQGGSVPRREFTGQIIPSTPHLPGMGSLGSSSAVQEHODD 306

QY 1085 TQHNQODEMKNKASHQKFTFLDLNLSSECTRGOSSTKONTIDGCLPRDTADVDVPLSN 1144
 Db 307 TQHNQODEMKNKASHQKFTFLDLNLSSECTRGOSSTKONTIDGCLPRDTADVDVPLSN 366
 QY 1145 NSSLONILVESNSSKKEQTAVYEKETNATILREKAGTADGKKPISQWDLKKDYBNGE 1204
 Db 367 NSSLONILVESNSSKKEQTAVYEKETNATILREKAGTADGKKPISQWDLKKDYBNGE 426
 QY 1205 ROERKNNMDSIDYPAIRASISEISEAIKERGMNMLAVRIKDLERIVKHGGIDLEW 1264
 Db 427 ROERKNNMDSIDYPAIRASISEISEAIKERGMNMLAVRIKDLERIVKHGGIDLEW 486
 QY 1265 LRESPDRAKDYLLSIRGLGKSVCEVRLTLNLAFVDTNVGRIAVRMGVVPLQPIPE 1324
 Db 487 LRESPDRAKDYLLSIRGLGKSVCEVRLTLNLAFVDTNVGRIAVRMGVVPLQPIPE 546
 QY 1325 SIQLHLEL 1333
 Db 547 SIQLHLEL 555

RESULT 3

DB4781
 hypothetical protein At2g36490 [imported] - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 02-Feb-2001 #sequence_rev1sion 02-Feb-2001 #text_change 02-Feb-2001
 C:Accession: DB4781
 R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Beilto, M.I.; Town, C.D.; Fujii, C.Y.
 euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Ventler
 Nature 402, 761-768, 1999
 A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
 A:Reference number: A84420; MUID:20083487
 A:Accession: DB4781
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-1207 <STO>
 A:Cross-references: GB:AE02093; MID:94581149; PIDN:AMD24633.1; GSPDB:GN00139
 A:Genetics:
 A:Gene: At2g36490
 A:Map position: 2

Query Match 17.2%; Score 1559.5; DB 2; Length 1207;
 Best Local Similarity 28.5%; Pred. No. 1.1e-76;
 Matches 480; Conservative 199; Mismatches 366; Indels 641; Gaps 61;

QY 13 TEATPONDGSRQDVLFDLNTPOQKPSKRRKFRMPKVVVEGKPKRPAELPKVVE 72
 Db 65 TEEVESLSVSNVAE-QILKTP-EKPRKKHR--PKVREAKPKRPRAPRKSVTVD 120
 QY 73 GKPRKPKKATQKVK-SKETGSAKKNLKESATKPRANYGDMSSNKPETLSCRAL 131
 Db 121 GQESTPKRKTVRKKEVSKD-----QDATPVSSAAVETSTRPK--RLCRVL 167
 QY 132 NFEDLEN-----PGDAR-GDSESEIVONSSGANSFSEIRDAIGTNGSFQDSVQIDKTN 185
 Db 168 DFEADENENQNGSIRREAGMESALQEKODSGN-QELKCL-----LSAPSTPKRKR 219
 QY 186 GLGAMNAPLEVSMGNQDPK-----LSTGAKIARDQOPDLTLRN-----QOCQFP 229
 Db 220 SQGKRK-----GVQPKKNGSNLEEVDISMAQAARKRQGPCTCDMLSLGIQYDEQCY- 271
 QY 230 VATONTQFPMENOQANLQMKNOQLGFFPGNOQPMATIRNOQPCLAMGNOQPMYILGTPRP 289
 Db 272 -----QKMHMDYSPNL-----QOGGM--RYDAIC----- 293
 QY 290 ALVAGNOOLGAPQGNKRPILFNHOTCLPAGNOLYGSPTDMHQLVMSNGGQOHGLIRNOQ 349
 Db 294 -----SKVFGQOHNYSAF--HATCTSSISQLSANRY-----LTVERK 331
 QY 350 PGLIRGOQPCVPLIDQOAPTP-----KGFHLNOMVATSSSPGLRPHSOSQV 398
 Db 398 PGLIRGOQPCVPLIDQOAPTP-----KGFHLNOMVATSSSPGLRPHSOSQV 398

QY	1444	PSUREN-----CEPIIEPASPGEQCTEITSIEDIAYINEDDPDIETIKLTNEQFGMT	1497
	:		:
Db	1000	VQHSEPAKVKTCCEPIIEPASPPEETAESTADIEAFF-EDPEIPTIRLNMDAFSTN	1058
	:		:
QY	1498	LREHMERNMELQGDMSKALVALHPTTTSIPPKLKNISRLTEHQVYELDPSHRLLDGM	1557
	:::		:
Db	1059	LKXIMEHNKLQDGNMSSALVALTAETASLPMPKLKNTSQLTEHRVYELDPDEPLLAQL	1118
	:::		:
QY	1558	KREPDDPPYLIAIWTPCETANSAPQBCKGCGASKCMCFDETCSCNSILREANSOTV	1617
	:		:
Db	1119	EKREPDDPCSyllAIWTPCETADSQPVSVCIFOANGMLCDCEITFCSCNSIKETRISOIV	1178
QY	1618	RGTLLI	1623
	:		:
Db	1179	RGTILV	1184

T48454

C; species:

C; Accessio

ewes, H.W.

A; References

A;Status: [

A; Residues

A; Expert Interview
C; Content Analysis

A: Map post
A: Introns:

SECRET

Query Mat

Matches

Qy 1496 R

Db 1 K

Qy 1556 C

Db 61 C

9716 T

121 99

0107 20

5 Tot 22

RESULTS

hypothetika

C; Date: 23-

R; Bevan, M.

A; Reference

A; Molecule

THIS PAGE BLANK (USPTO)